Four Essays in Microeconometrics

Dissertation
for the Faculty of Economics, Business Administration and Information Technology of the University of Zurich

to achieve the title of
Doctor of Philosophy
in Economics

presented by

Gregori Bätschmann
from Jonschwil SG

approved in April 2013 at the request of

Prof. Dr. Rainer Winkelmann
Prof. Dr. Michael Wolf

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorises the printing of this Doctoral Thesis, without thereby giving any opinion on the views contained therein.

Zurich, April 3, 2013

Chairman of the Doctoral Committee: Prof. Dr. Dieter Pfaff

Acknowledgements

In writing this dissertation, I benefited from the support and inspiration of many people. Without their help, I would not have been able to complete this project.

First of all, I like to thank my supervisor Prof. Rainer Winkelmann for whom I worked at the Chair for Statistics and Empirical Economic Research at the University of Zurich. His encouragement and guidance during the last years has been very important to me. I would also like to thank Prof. Michael Wolf for co-advising my thesis.

I had many stimulating discussions with Kevin Staub, Raphael Studer, Timo Boppart, Johannes Kunz, Stefan Boes, and Florian Schaffner. I am very grateful for these conversations, which deepened my understanding of econometrics and other topics.

Finally, I want to thank my parents, my sister and Barbara for their encouragement.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The four essays of this dissertation deal with different topics in microeconometrics, including model building, identification, and estimation. This section first offers a brief summary of each chapter. Next, I discuss some overarching themes of the essays.

1.1 Key results

Chapter 2, which is jointly written with Kevin Staub and Rainer Winkelmann, reconsiders existing estimators for the panel data fixed effects ordered logit model and studies their small sample properties in a series of Monte Carlo simulations. The main finding is that some estimators used in the literature are inconsistent. The empirical relevance of the ordered logit model is illustrated in an application, which studies the effect of unemployment on happiness. Different econometric approaches are used to avoid a bias in the estimated effect of unemployment on happiness if the employment status is correlated with unobserved variables. The results confirm the large negative effect of unemployment found in the previous literature.

Chapter 3 proposes an approach to estimate the so called "thresholds" in the fixed effects ordered logit model. Knowing the thresholds is helpful for interpreting the regression coefficient, and enables statements about the effect of a changing covariate on the observed ordered variable – and not only on the latent variable as in Chapter 2. The empirical relevance of the new estimation procedure is again illustrated by considering the effect of unemployment on happiness. The main finding is that the observed ordered variable of individuals having a low satisfaction score is more strongly affected by a change in the unemployment status compared to individuals with a low satisfaction score.

Chapter 4 studies the evolution of life satisfaction over the life course in Germany. It clarifies the causal interpretation of the econometric model by discussing the choice of control variables and the underidentification between age, cohort, and time effects. The empirical part analyzes the distribution of happiness over the life course at the aggregated level, at the subgroup level, and at the individual level. There are several important findings. First, on average, life satisfaction is slowly decreasing up to age fifty-five followed by a hump shape with a maximum at seventy. Secondly, the analysis at the lower levels suggests that people differ in their life satisfaction trends, whereas the hump shape after age fifty-five is robust. Third, no significant differences between men and women are found. In contrast, education groups differ in their trends: Better educated people become happier over the life cycle, while life satisfaction decreases for less educated people.

Chapter 5, finally, which is jointly written with Rainer Winkelmann, provides a new explanation of extra zeros in count data models, related to the underlying stochastic process that generates events. The process has two rates, a lower rate until the first event, and a higher one thereafter. We derive the corresponding distribution of the number of events during a fixed period and extend it to account for observed and unobserved heterogeneity. The new stochastic hurdle model allows to address the effect of exposure time in a theory-consistent way, and it overcomes limitations of previous decompositions into extensive and intensive margin effects. The new approach is empirically illustrated by an analysis of the effect of a health care reform on the individual number of doctor visits in Germany.

1.2 Overarching themes

Here I discuss four overarching themes of my dissertation. In particular, I want to emphasize the commonalities of the data, the identification issues, the type of the dependent variable, and the relevance of the studied topics for empirical research. I consider each of the commonalities in turn.

First, all four essays use data from the German Socio Economic Panel (SOEP). I use this commonality as starting point to discuss the papers by focusing on the data. The SOEP is a longitudinal panel dataset based on German households. Data collection started in 1984 and people are repeatedly interviewed on a yearly basis. The SOEP is a microdata survey. This brings up the question to which extend it can be viewed as a random sample from an underlying population. All my essays focus on relationships between variables surveyed in the SOEP, such as unemployment and life satisfaction. Thus representativeness for the underlying German population is not as crucial as if one is interested in statistics of specific variables, such as the average happiness score in Germany. However, sample selection or the problem of endogeneity can also be seen as a lack of a representative sample, namely from the counterfactual outcomes (e.g. Hollande, 1986). The issue of sample selection emerges especially in the contributions about life satisfaction. Specifically in the two chapters dealing with the relationship between unemployment and happiness, the concern is that a third factor might affect both variables. This would imply that regressing life satisfaction on unemployment status would not estimate the causal effect. However, the issue of sample selection would not be present if experimental data instead of observational data were used.

The use of observational data to answer a causal question can be defended in two ways. First, one places the researcher in a purely data consuming position, where only observational data are available. Alternatively, one could argue that running an experiment is just impossible, which is often the case in a macroeconomic context. Based on these arguments, one can either ignore the question because it seems impossible to obtain assured knowledge or work with observational data. The effect of unemployment on life satisfaction seems to be one of these questions where running an experiment is at least difficult. The question, however, seems to be too important for not being addressed.

In Chapter 4, the issue of representativeness is even more complex since it is not possible to solve it by running a simple experiment. The problem is that people interviewed several times report a lower satisfaction score compared to people surveyed the first time, implying an effect of interviewing on the answer. This suggests interviewing each person only once. However, this makes it impossible to examine whether or not the u-shape relationship between age and life satisfaction, found at the aggregated level by several authors (e.g. Blanchflower and Oswald, 2008), is also present at the individual level, or if it is just a result of mixing different non u-shaped forms. The main reason why repeated measurement affects the outcome so strongly is the subjective character of self-reported life satisfaction. In my opinion, replacing the subjective statement by an objective measurement is not a viable option.

Second, drawing upon the previous discussion, I want to emphasize the common issue of identification. The Chapters 2 and 3 show, for the fixed effects ordered logit model, how to estimate the thresholds and the effect of covariates on the latent index. However, even knowing these standard parameters of the model does not allow to identify important statistics like the average marginal effect, because the distribution of the individual fixed effects is unknown (see Honoré and Tamer (2006) for partial identification in the binary logit model). The topic of underidentification in Chapter 4 was already touched by mentioning the issue with repeated measurements. In fact, another identification problem is caused by the linear dependence between age, time and birth cohort. This hinders to estimate the linear effect of all three variables (only a combination of the linear effects is identified). In Chapter 5, the issue of underidentification emerges in the context of decomposing the effect of a policy reform into extensive and intensive margin effects, thus into effects due to the changed fraction of zeros and the shifts in the strict positive part of the distribution.

Third, I want to highlight that the dependent variables in the essays have limited support. In Chapters 2 to 4, the dependent variable is self-reported life satisfaction measured on a discrete scale ranging from 0 to 10. However, the variable is treated differently in the three essays. While Chapter 2 and 3 consider the happiness variable as an ordered one (thus use only the order of the eleven categories and not the numeric value of them) Chapter 4 focuses on the conditional expectation function estimated by ordinary least squares (OLS). Neglecting the ordered scale can be defended in two ways: First by referring to the empirical literature stating that the regression model, in contrast to the inclusion of individual fixed effects, does usually not affect the qualitative results of happiness regressions (Ferrer-i-Carbonell and Freijters, 2004). Second, by pointing out the simplicity of OLS. Linear regression focuses on the expectation function and enables to present the findings in a clear and understandable way. In Chapter 4, the whole distribution of the dependent variable over the life course is studied and a weak common pattern for all categories is found. If the variable is treated as numeric, this common pattern is mirrored in the u-shape followed by a hump shape of the conditional expectation function. Another possibility to summarize the shared shape is the conditional median. However, it is here not sensitive enough to summarize the small distributional changes over the life course. Chapter 5 treats doctoral visits as a count variable and thus also as a limited dependent variable. In contrast to an ordered dependent variable, the expectation function exists and can be estimated by OLS. However, assuming a linear relationship between the covariates and doctoral visits is difficult to defend, because of the limited support of the dependent variable. A nonlinear model seems to be more appropriate. The advantage of a count data model, which implies a nonlinear relationship, compared to semi or non-parametric models is that it is simpler to interpret and facilitates extrapolation.

Fourth, and finally, the econometric problems studied in this thesis are directly linked and motivated by widespread empirical practice, in contrast to innovations motivated by theoretical concerns or concrete applications. The ordered logit model is one of the workhorse models to analyze ordered dependent variables, and including fixed effects an obvious extension. Knowing how to estimate the parameters is therefore relevant beyond this particular application. The econometric contribution of the Chapter 4 is to describe the identification problem if the model includes covariates which are linear dependent, and to stress that finding a stable relationship between variables on the aggregated level does not ensure that the same relation does also hold on the individual level. Both problems are also present in other contexts. Chapter 5 is also directly motivated by a widespread concern in empirical research, the frequent zeros in empirical count data distributions. The suggestion is to use a model with two different rates, a lower initial until the occurrence of the first count and a higher thereafter. It is as simple as the standard models for count data with many zeros, the zero inflated and the hurdle model, but has a representation in terms of a plausible latent count process. An advantage of having such a latent process is the possibility to include exposure time in a theory consistent way.

References

Holland, P.W. (1986): “Statistics and causal inference”, *Journal of the American Statistical Association*, **81**, 945-970.

Honoré, B. E., and Tamer, E. (2006): “Bounds on Parameters in Panel Dynamic Discrete Choice Models”, *Econometrica*, **74**, 611-629.

Ferrer-i-Carbonell, A., and P. Frijters (2004): “How important is methodology for the estimates of the determinants of happiness?”, *The Economic Journal*, **114**, 641-659.

Blanchflower, D. G., and Oswald, A. J. (2008): “Is Wellbeing U-shaped over the Life Cycle”, *Social Science and Medicine*, **66**, 1733-1749.

Chapter 2

Reconsidering panel data methods for ordered response variables – with an application to the effect of unemployment on happiness

This chapter is joint work with Rainer Winkelmann and Kevin Staub. An earlier version with the title "Reconsidering the analysis of longitudinal happiness data – with an application to the effect of unemployment" was published as Working Paper No. 4 in the *Working Paper Series* of the Department of Economics, University of Zurich.

Acknowledgements: We thank three anonymous referees, Thierry Magnac, Paul Frijters, Arie Kapteyn as well as participants of the Engelberg Workshop in Labor Economics, the North American Summer Meetings of the Econometric Society and the 17th International Panel Data Conference for very valuable comments on an earlier version.

2.1 Introduction

When estimating happiness equations, or analyzing determinants of job satisfaction or self-assessed health, researchers are often concerned about unobserved heterogeneity. Such heterogeneity can result from omitted variables or from subjective differences in anchoring of responses on the ordered scale. If unaccounted for, heterogeneity will generally bias the estimated effects. Panel data offer a promising solution to this problem as they provide more information that can be used to construct consistent estimators as long as the unobserved heterogeneity is time-invariant.

Unfortunately, there is no consensus in the past literature on how to implement a fixed effects estimator for the ordered logit model. In order to avoid the incidental parameter problem, all versions of the fixed effects ordered logit estimator rely on conditional logit estimation of a dichotomized response (Chamberlain, 1980). In an early application, Winkelmann and Winkelmann (1998) used a single dichotomization at a constant value for all individuals in order to estimate the effect of unemployment on life satisfaction. Das and van Soest (1999) suggested to increase efficiency by combining estimators for different dichotomizations using a two-step minimum distance estimator. Ferrer-i-Carbonell and Frijters (2004) by contrast proposed to use an individual-specific dichotomization in order to minimize the variance of the estimator. The minimum distance (MD) approach has to the best of our knowledge not been used in applied work since, at least not in the areas of happiness, health and job satisfaction, whereas applications of the Ferrer-i-Carbonell and Frijters (FF) estimator are quite frequent and include Frijters, Haisken-DeNew and Shields (2004a, 2004b, 2005), Kassenboehmer and Haisken-DeNew (2009), Booth and van Ours (2008), D'Addio, Eriksson and Frijters (2007), Schmitz (2011) and Jones and Schurer (2011).

The contribution of this paper is both substantive and methodological. On the substantive side, we provide new evidence on the causal effect of unemployment on happiness. We use panel data on working-age men in Germany for the period 1991-2009 and estimate the parameters using the consistent and efficient MD estimator, plus a modified version of it that is simpler to implement. In addition we account for potential additional endogeneity due to correlated shocks by using plant closure as an instrument for unemployment. In order to construct an IV estimator in this non-linear set-up, we adapt the special regressor approach developed by Honoré and Lewbel (2002) to the ordered logit model with fixed effects. Neither the MD nor the special regressors approach have been used before in this context. The results corroborate the earlier findings in the literature. The adverse effect of unemployment on life satisfaction is large.

On the methodological side, we show in this paper that the FF estimator is inconsistent. We discuss different approaches which use all the available information and are consistent. The first of these is the aforementioned two-step minimum distance estimator. As an alternative to the MD estimator, we investigate generalized methods of moments (GMM) and empirical likelihood (EL) estimators. We also discuss another consistent estimator that has been introduced in the statistics literature by Mukherjee at al. (2008) but not been applied in econometric studies to date. For reasons that become apparent when we introduce the estimator in detail, we refer to it as "blow-up and cluster" (BUC) estimator. The BUC estimator is simple to implement but asymptotically inefficient relative to MD, GMM and EL. However, it avoids some small sample problems that can limit the usefulness of these estimators in applied work.

The paper proceeds as follows. Section 2.2 reviews the different estimators for the fixed effects ordered logit model. Section 2.3 reports results from a Monte Carlo study in order to compare the performance of the various estimators as a function of sample size (number of individuals and number of time periods) as well as number of ordered categories. The analysis of the effect of unemployment on life satisfaction, using data from the German Socio-economic Panel, follows in Section 2.4. Section 2.5 concludes.

2.2 Econometric methods

2.2.1 The fixed effects ordered logit model

The fixed effects ordered logit model relates the latent variable y_{it}^* for individual i at time t to a linear index of observable characteristics x_{it} and unobservable characteristics α_i and ε_{it}:

$$y_{it}^* = x_{it}'\beta + \alpha_i + \varepsilon_{it}, \quad i = 1, \ldots, N \quad t = 1, \ldots, T \ . \tag{2.1}$$

The time-invariant part of the unobservables, α_i, may or may not depend on x_{it}. One can either make an assumption regarding the distribution of α_i (or the joint distribution of α_i and x_{it}), or else treat α_i as a fixed effect. This paper considers estimation under the fixed effects approach. The observed ordered variable y_{it} is tied to the latent variable by the observation rule:

$$y_{it} = k \quad \text{if} \quad \tau_{ik} < y_{it}^* \leq \tau_{ik+1}, \quad k = 1, \ldots, K \ , \tag{2.2}$$

where individual-specific thresholds τ_i are increasing ($\tau_{ik} \leq \tau_{ik+1} \ \forall k$), $\tau_{i1} = -\infty$, and $\tau_{iK+1} = \infty$. Moreover, the fixed effects ordered logit model assumes that ε_{it} are i.i.d distributed with logistic cdf

$$F(\varepsilon_{it}|x_{it}, \alpha_i) = F(\varepsilon_{it}) = \frac{1}{1 + \exp(-\varepsilon_{it})} \equiv \Lambda(\varepsilon_{it}) \ . \tag{2.3}$$

Hence, the probability of observing outcome k for individual i at time t is given by

$$\Pr(y_{it} = k|x_{it}, \alpha_i) = \Lambda(\tau_{ik+1} - x_{it}'\beta - \alpha_i) - \Lambda(\tau_{ik} - x_{it}'\beta - \alpha_i) \ , \tag{2.4}$$

which depends not only on β and x_{it}, but also on α_i, τ_{ik} and τ_{ik+1}. It is clear from (3.4) that only $\tau_{ik} - \alpha_i \equiv \alpha_{ik}$ is identified. Moreover, under fixed-T asymptotics, α_{ik} cannot be estimated consistently by maximum likelihood, due to the incidental parameter problem (see, for instance, Lancaster, 2000), and neither can β. In short panels, the resulting bias in $\hat{\beta}$ can be substantial (Greene, 2004). Instead, a consistent estimator of β is obtained from collapsing y_{it} into a binary variable and then applying conditional maximum likelihood (CML) estimation (Chamberlain, 1980).

The CML estimator is well known, but we present it nevertheless in some detail in order to fix notation. Let d_{it}^k denote the binary dependent variable that results from dichotomizing the ordered variable at the cutoff point k: $d_{it}^k = \mathbb{1}(y_{it} \geq k)$. k can be any integer between 2 and K. By construction, $\Pr(d_{it}^k = 0) = \Pr(y_{it} < k) = \Lambda(\tau_{ik+1} - x_{it}'\beta - \alpha_i)$, and $\Pr(d_{it}^k = 1) = 1 - \Lambda(\tau_{ik+1} - x_{it}'\beta - \alpha_i)$. Now consider the joint probability of observing $d_i^k = (d_{i1}^k, \ldots, d_{iT}^k)' = (j_{i1}, \ldots, j_{iT})'$ with $j_{it} \in \{0,1\}$. The sum of all the individual outcomes over time is a sufficient statistic for α_i as

$$\mathcal{P}_i^k(\beta) \equiv \Pr\left(d_i^k = j_i \,\middle|\, \sum_{t=1}^{T} d_{it}^k = g_i\right) = \frac{\exp(j_i' x_i \beta)}{\sum_{j \in B_i} \exp(j' x_i \beta)} \tag{2.5}$$

does not depend on α_i and the thresholds. In (2.5), $j_i = (j_{i1}, \ldots, j_{iT})$, x_i is the $(T \times L)$-matrix with tth row equal to x_{it}, L is the number of regressors and $g_i = \sum_{t=1}^{T} j_{it}$. The sum in the denominator goes over all vectors j which are elements of the set B_i

$$B_i = \left\{ j \in \{0,1\}^T \,\middle|\, \sum_{t=1}^{T} j_t = g_i \right\},$$

i.e., over all possible vectors of length T which have as many elements equal to 1 as the actual outcome of individual i, g_i. The number of j-vectors in B_i is T choose g_i. Chamberlain (1980) showed that maximizing the conditional log likelihood

$$LL^k(b) = \sum_{i=1}^{N} \log \mathcal{P}_i^k(b) \tag{2.6}$$

gives a consistent estimator for β, denoted by $\hat{\beta}^k$ and henceforth referred to as *Chamberlain estimator* (see also Andersen, 1970). The first order conditions are $\sum_i s_i^k(b) = 0$ where

$$s_i^k(b) = \frac{\partial \log \mathcal{P}_i^k(b)}{\partial b} = x_i' \left(d_i^k - \sum_{j \in B_i} j \frac{\exp(j' x_i b)}{\sum_{l \in B_i} \exp(l' x_i b)} \right) \tag{2.7}$$

and the asymptotic variance of $\hat{\beta}^k$ is given by

$$\text{AVar}(\hat{\beta}^k) = -\text{E}(H_i^k(\beta))^{-1} = \text{E}(s_i^k(\beta) s_i^k(\beta)')^{-1}, \tag{2.8}$$

which can be estimated by averaging over individual Hessians

$$H_i^k(b) = \frac{\partial^2 \log \mathcal{P}_i^k(b)}{(\partial b)(\partial b)'} = -\sum_{j \in B_i} \frac{\exp(j'x_i b)}{\sum_{l \in B_i} \exp(l'x_i b)} \times$$
$$\left(x_i'j - \sum_{m \in B_i} \frac{\exp(m'x_i b)}{\sum_{l \in B_i} \exp(l'x_i b)} m'x_i \right) \left(x_i'j - \sum_{m \in B_i} \frac{\exp(m'x_i b)}{\sum_{l \in B_i} \exp(l'x_i b)} m'x_i \right)' . \quad (2.9)$$

An important property of the Chamberlain estimator is that individuals with constant d_{it}^k do not contribute to the conditional likelihood function, since $\Pr(d_{it}^k = 1 | \sum_{t=1}^T d_{it}^k = T) = \Pr(d_{it}^k = 0 | \sum_{t=1}^T d_{it}^k = 0) = 1$. However, the ordered dependent variable can be dichotomized at different cutoff points resulting in several consistent Chamberlain estimators. With K ordered outcomes, there are $K-1$ such estimators, and they employ information from different groups of individuals, depending on who crosses the cutoff and thus has variation in the dichotomized variable. For each individual there is at least one $k = \tilde{k}$ such that $d_{it}^{\tilde{k}}$ is not constant, unless $y_{i1} = \ldots = y_{iT}$. This feature is exploited by the individual-specific cutoff estimators discussed in the next section.

2.2.2 Individual-specific cutoff points

Ferrer-i-Carbonell and Frijters (2004) suggested to use a single but distinct, in some sense "optimal", cutoff point for each individual. A compact way of writing the FF estimator is by way of a weighed conditional log-likelihood function

$$LL^{FF}(b) = \sum_{i=1}^{N} \sum_{k=2}^{K} w_i^k \log \mathcal{P}_i^k(b) , \quad (2.10)$$

where $\mathcal{P}_i^k(b)$ is defined as in (2.5), $w_i^k = 0, 1$ and $\sum_{k=2}^K w_i^k = 1$. This objective function is maximized with respect to b, conditional on the individual's weight vector w_i^k, $k = 2, \ldots, K$. The crucial question is where to dichotomize the dependent variable or, equivalently, which w_i^k to set to one. The FF approach is to calculate for every individual all Hessian matrices under different cutoff points and then to choose the cutoff with the smallest Hessian:

$$w_i^k = 1, \text{ if } k = \underset{k}{\operatorname{argmin}}\, H_i^k(\beta).$$

In practice, the Hessians are evaluated at $\hat{\beta}$, a preliminary consistent estimator. By choosing the cutoff point leading to the smallest Hessian, this rule should yield a fixed effects ordered logit estimator with the smallest inverse of minus the sum of the Hessians, and thus minimal variance. Other, simpler rules for choosing w_i^k have been used in the literature, trading efficiency for computational convenience. In fact, the standard way in which this estimator is implemented in applications is by choosing the individual mean of the dependent variable as dichotomizing cutoff point. Another possibility is to dichotomize at the median.

The key point is that these procedures determine the dichotomizing cutoff point endogenously, since it depends on y_i. This is problematic and leads to an inconsistent estimator. In order to provide some intuition for the inconsistency, consider the mean-cutoff estimator as an example. In that estimator, it is easily seen that the cutoff is endogenous since $d_{it}^{\mathrm{Mn}} = 1$ if and only if $y_{it} \geq T^{-1} \sum_t y_{it}$. Thus, y_{it} itself is part of the cutoff, and the probability $\Pr(d_{it}^{\mathrm{Mn}} = 1)$ can be written as

$$\Pr(d_{it}^{\mathrm{Mn}} = 1) = \Pr\left(y_{it} \geq \frac{1}{T} \sum_t y_{it}\right) = \Pr\left(y_{it} \geq \frac{1}{T-1} \sum_{s \neq t} y_{is}\right) .$$

The expression after the first equality makes it clear that for any t, y_{it} is on both sides of the inequality sign. Solving for y_{it} shows that the probability $\Pr(d_{it}^{Mn} = 1)$ is equal to the probability that the outcome in t is greater than the average outcome in the remaining periods. In general, this is a different value for every period, and the implicit within-individual correlation between y_{it} and the time-varying cutoff is negative. With endogenous cutoffs, the score of these estimators does not converge to zero at the true parameter. A working paper version of this paper, Baetschmann, Staub and Winkelmann (2010), gives a formal proof of inconsistency. In Section 2.3, we provide quantitative information on the magnitude of the bias in a number of Monte Carlo simulations.

2.2.3 Consistent and efficient estimators

The estimators discussed so far use only one dichotomization per individual to estimate β. This implies that they do not use all information contained in the variation of the dependent variable, and alternative approaches can provide efficiency gains. A first possibility is to separately calculate all Chamberlain estimators and then combine them in a second step using minimum distance estimation (Das and van Soest, 1999). The second approach estimates β based on the sum of the likelihood functions of all the different Chamberlain estimators. This method was used for instance by Mukherjee et al. (2008). The third approach is to combine the moment restrictions implied by the model and use them in a GMM framework to estimate β.

Minimum Distance Estimation

Since every Chamberlain estimator $\hat{\beta}^k$ is a consistent estimator of β, so is any weighted average of them. The efficient combination can be obtained by minimum distance (MD) estimation. Specifically, let M be a matrix of $K-1$ stacked L-dimensional identity matrices, and $\tilde{\beta}$ the $(K-1) \cdot L \times 1$ vector containing the $K-1$ Chamberlain estimators. The MD estimator is given by:

$$\hat{\beta}^{MD} = \underset{b}{\operatorname{argmin}}\, (\tilde{\beta} - Mb)' \operatorname{Var}(\tilde{\beta})^{-1} (\tilde{\beta} - Mb), \tag{2.11}$$

where $\operatorname{Var}(\tilde{\beta})$ is the variance-covariance matrix of the stacked Chamberlain estimators (Das and van Soest, 1999). The solution to (2.11) is

$$\hat{\beta}^{MD} = \Big(M'\operatorname{Var}(\tilde{\beta})^{-1}M\Big)^{-1} M'\operatorname{Var}(\tilde{\beta})^{-1}\tilde{\beta}\,,$$

showing that the MDE is a matrix weighted average of the Chamberlain estimators. The asymptotic variance (i.e., the limiting variance of $\sqrt{n}(\hat{\beta}^{MD} - \beta)$) is

$$\begin{aligned}
\operatorname{AVar}(\hat{\beta}^{MD}) &= (M'\operatorname{AVar}(\tilde{\beta})^{-1}M)^{-1} \\
&= (\mathrm{E}(H_i(\beta))'\mathrm{E}(s_i(\beta)s_i(\beta)')^{-1}\mathrm{E}(H_i(\beta)))^{-1},
\end{aligned}$$

where $s_i(\beta)$ denotes individual i's stacked Chamberlain scores evaluated at β, and $H_i(\beta)$ the stacked Hessians (see Appendix A).

Restricted CML estimation

Alternatively, the information associated with the different cutoffs can be combined in a single likelihood function, leading to a one-step estimator of β. The sample (quasi-) log-likelihood function of this restricted CML estimator is

$$LL^{BUC}(b) = \sum_{k=2}^{K} LL^k(b)\,, \tag{2.12}$$

where $LL^k(b)$ is defined as in (2.6), and $\hat{\beta}^{BUC}$ is the estimator that maximizes (2.12) and thus imposes the restriction that $\hat{\beta}^2 = \ldots = \hat{\beta}^K$. Such an estimator has been suggested by Mukherjee et al. (2008). We refer to it is *blow-up and cluster* (BUC) because that describes the way of implementing this estimator using conditional maximum likelihood estimation: Replace every observation in the sample by $K-1$ copies of itself ("blow-up" the sample size), and dichotomize each of the $K-1$ copies of the individual at a different cutoff point. The BUC estimates are obtained by CML estimation using the entire sample. The standard errors need to be clustered at the individual level since observations are dependent by construction.

It is straightforward to see that this approach leads to a consistent estimator. The score of the BUC log-likelihood function equals the sum of the scores of the Chamberlain estimators. Since these estimators are consistent, their scores converge to zero in probability at the true parameter. It follows that the probability limit of the score of the restricted CML estimator is zero as well:

$$\text{plim} \sum_{k=2}^{K} \frac{1}{N} \sum_{i=1}^{N} s_i^k(\beta) = \text{plim} \frac{1}{N} \sum_{i} s_i^2(\beta) + \ldots + \text{plim} \frac{1}{N} \sum_{i} s_i^K(\beta) = 0 \tag{2.13}$$

which, together with the concavity of the objective function, implies that $\hat{\beta}^{BUC}$ converges to β.

Since some individuals contribute to several terms in the log-likelihood this creates dependence between these terms, invalidating the usual estimate of the estimator variance based on the information matrix equality. Instead, a cluster-robust variance estimator which allows for arbitrary correlation within the various contributions of any individual should be used. The formula for the variance can be found in the next section, where it is shown that the BUC estimator can be written as an inefficient GMM estimator. The main difference between MD and BUC estimation is the weighting: by simply summing over the log-likelihood contributions, the BUC estimator implies different weights of the Chamberlain estimators than the variance-based weights used by MDE.

GMM and Empirical Likelihood

A third approach for achieving efficiency gains over the simple Chamberlain estimator combines the moment conditions implied by the model under the different dichotomizations. With L explanatory variables, each dichotomization leads to L zero-expected score moment conditions. This gives $(K-1)\cdot L$ restrictions in total. Since only L parameters are estimated, the system is over-identified. The generalized method of moment (GMM) estimator with weighting matrix W is

$$\hat{\beta}^{GMM} = \underset{b}{\operatorname{argmin}}\, s(b)'Ws(b), \tag{2.14}$$

where $s(b)' = \frac{1}{N}\sum_{i=1}^{N}(s_i^{2'}(b), \ldots, s_i^{K'}(b))$. The first order conditions of the GMM estimator with weighting matrix W are given by

$$\frac{\partial s(\hat{\beta}^{GMM})}{\partial \hat{\beta}^{GMM'}} W s(\hat{\beta}^{GMM}) = H(\hat{\beta}^{GMM})' W s(\hat{\beta}^{GMM}) = 0. \tag{2.15}$$

where $H(\hat{\beta}^{GMM})$ denotes the matrix of stacked Hessians of the single Chamberlain estimators evaluated at $\hat{\beta}^{GMM}$: $H(b)' = (H^2(b), \ldots, H^k(b))$. The efficient GMM estimator uses the inverse of the variance of the moment conditions as weighting matrix: $W = \mathrm{E}(s(\beta)s(\beta)')^{-1}$.

The asymptotic variance of efficient GMM is

$$\begin{aligned}\text{AVar}(\hat{\beta}^{GMM}) &= \left[\text{E}\left(\frac{\partial s_i(\beta)}{\partial \beta'}\right)' (\text{E}(s_i(\beta)s_i(\beta)')^{-1}\text{E}\left(\frac{\partial s_i(\beta)}{\partial \beta'}\right)\right]^{-1} \\ &= (\text{E}(H_i(\beta))'\text{E}(s_i(\beta)s_i(\beta)')^{-1}\text{E}(H_i(\beta)))^{-1}. \end{aligned} \tag{2.16}$$

It equals the asymptotic variance of the minimum distance estimator. The form of the first order conditions, equation (2.15), implies a GMM representation of the BUC estimator: Setting the weighting matrix to a block diagonal matrix with the inverse of the Chamberlain Hessians on the diagonal yields the first order conditions of the BUC estimator. Since this matrix is not equal to the weighting matrix of the efficient GMM estimator, the BUC estimator has in general a larger variance than the MD and GMM estimators. Using standard GMM results, the asymptotic variance of the BUC estimator is:

$$\begin{aligned}\text{AVar}(\hat{\beta}^{BUC}) &= (\text{H}_i'W^B\text{H}_i)^{-1}(\text{H}_i'W^B\text{S}_iW^B\text{H}_i)(\text{H}_i'W^B\text{H}_i)^{-1} \\ &= \left(\sum_{k=2}^{K}\text{E}(H_i^k(\beta))\right)^{-1}\left(\sum_{k=2}^{K}\sum_{l=2}^{K}\text{E}(s_i^k(\beta)s_i^l(\beta))\right)\left(\sum_{k=2}^{K}\text{E}(H_i^k(\beta))\right)^{-1}, \end{aligned} \tag{2.17}$$

with $\text{H}_i = \text{E}(H_i(\beta))$, $\text{S}_i = \text{E}(s_i(\beta)s_i(\beta)')$, and W^B denoting the described weighting matrix. The second equality follows since $W^B\text{H}_i = M$, thus a matrix of $K-1$ stacked L-dimensioned identity matrices. An estimate of expression (2.17) can be used to construct optimal weights for a weighted version of the BUC estimator.

As an alternative to GMM, the empirical likelihood (EL) estimator works directly with moment conditions as well. It has the identical asymptotic distribution as the efficient GMM estimator. However, EL estimators usually have better small sample properties (see e.g. Kitamura, 2006). In our set-up, the EL estimator is the result to the following optimization problem:

$$\max_{p,b}\sum_{i=1}^{N}\log(p_i), \quad \text{subject to } \sum_{i=1}^{N}p_i = 1 \text{ and } \sum_{i=1}^{N}s_i(b)p_i = 0. \tag{2.18}$$

The vector $s_i(b)$ is the vector of stacked Chamberlain scores for individual i. p_i denotes the probability of observing individual i's variable realizations. The interest is only in b, whereas p is treated as an auxiliary parameter vector.

2.3 Monte Carlo study

This section compares bias, precision, and overall robustness of the various estimators of the fixed effects ordered logit model in small samples using Monte Carlo simulations. First, although all estimators may suffer from bias in small and moderately sized samples due to the non-linearity of the objective functions, this bias, if any, should be minor compared to the bias from inconsistency due to endogenously chosen cutoffs by the FF estimators. Second, while MD, GMM and EL are more efficient than Chamberlain and BUC, this is an asymptotic result that requires the use of optimal weights. In practice, the weights are unknown and need to be estimated from the data. This can be problematic if the sample size is small and there is a large number of categories, so that the number of individuals who cross a certain threshold is low. This situation is frequently encountered in applied research. In such cases the performance of the estimators may be poor, or even worse, empirical counterparts of some of the moments may not be defined due to a lack of observations.

It is therefore not clear, ex-ante, whether the efficient estimators dominate the simpler ones in finite sample settings. Anticipating our results, we find that the estimator which suffers the least from such problems is the BUC estimator. It is approximately unbiased and the efficiency loss relative to the optimal estimators is very modest in our simulations. These facts make the BUC estimator an attractive option.

2.3.1 Experimental design

The data generating process (DGP) for the latent variable is

$$y^*_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + \alpha_i + \varepsilon_{it},\ i = 1, \ldots, N,\ t = 1, \ldots, T\ ,$$

where we set $\beta_1 = 1$, $\beta_2 = 1$. The continuous regressor x_1 is normally distributed $N(0, 0.5)$. ε_{it} has a standard logistic distribution. x_2 is a binary regressor that is correlated with α_i. Specifically, with probability 0.5, $x_2 = 0$ and $\alpha \sim N(0, 0.5)$ and with probability 0.5, $x_2 \sim Bernoulli(0.5)$ and $\alpha \sim N(1, 0.5)$. Thus, α is uncorrelated with x_1 and correlated ($\rho \approx 0.4$) with x_2. The observed ordered response variable y is obtained from the threshold mechanism (2.2). The number of categories K is equal to 5. In the first part of the DGP (where $\alpha \sim N(0, 0.5)$), all threshold are equal to zero. Thus, only outcomes $y = 1$ and $y = 5$ are observed. In the second part, the thresholds are chosen such that y follows a discrete uniform distribution. The baseline DGP is a balanced panel of N=500 individuals observed for T=3 periods. In a second step, the DGP is modified by increasing N, T and K.

2.3.2 Results

Table 2.1 contains results for the Monte Carlo simulations, based on 1,000 replications of each DGP. Columns with heading $\hat{\beta}_1$ show means of estimated coefficients corresponding to x_1, and columns labeled $\hat{\beta}_2$ show means for those corresponding to x_2. The numbers in parentheses are the standard deviations of the estimates. The first four columns provide the results of the baseline DGP: $N = 500$, $T = 3$, and $K = 5$; columns five to eight the results for a scenario where the number of individuals is increased to 1,000; columns nine to twelve the results for twice the number of time periods; and the last four columns show the results with ten instead of five categories. Each row of the table refers to a different estimator: the top four rows display results for Chamberlain estimators with cutoff points 2 to 5, followed by the three estimators with endogenous cutoffs. The last four rows contain the results of the four procedures combining the information of different cutoffs.

The simple Chamberlain estimators perform well in the simulations. They appear not to suffer from small sample bias and the estimation procedure always converge. However, they have a higher variance compared to the other proposed estimators. The efficiency loss, as one would expect, gets larger as the number of categories increases.

The estimators with endogenous cutoffs, in contrast, are clearly distorted. In the baseline scenario, the order of the biases is between 5% and 14%. The margin of error at the 99% level is always smaller than 0.02 and the distortion is therefore substantial. The bias is larger for the original FF estimator than for the mean or median cutoff estimators. To confirm the hypothesis that the deviation of the estimator's mean from the true parameter is caused by the procedure's inconsistency and not just a result of the small sample, we doubled the number of independent observations. The deviation is the same and illustrates therefore that these estimators are inconsistent.

Expanding the number of time periods, in contrast, reduces the distortion. The reason for the bias is not the use of individual-specific cutoffs *per se*, but the dependence of these cutoffs on y_i. Increasing the number of time periods decreases the dependence between cutoffs and realized error terms, and leads therefore to less biased estimators. On the other hand, the size of the bias is exacerbated by adding categories. This is to be expected since all estimators degenerate to the same consistent Chamberlain estimator if the number of categories shrinks to two. For example, for 10 categories, a standard number in research on job satisfaction and happiness, the mean of the FF estimator for β_2 is 0.80, well below the true value of 1.

Regarding the estimators which combine the available information of the Chamberlain estimators, it is noteworthy how well the BUC estimator performs. Although it is asymptotically less efficient, we find that the actual efficiency loss of the BUC estimator is small to negligible in our Monte Carlo simulations. Regarding distortions, neither the BUC nor the EL estimator seem to suffer from an observable small sample bias. The GMM and the MD estimator in contrast show signs of distortions. These are most accentuated if there are few observations and many categories. The bias for β_1^{MD}, for example, is around five percent in the scenario with 500 individuals, three time periods, and ten categories. The bias of the GMM estimator in this setting is about 6 percent. Another problem is that the GMM and EL estimators did not always converge, at least with our STATA implementation. It is known that this sort of convergence difficulties tends to be more pervasive with a higher number of explanatory variables. Thus, the BUC estimator can be a useful alternative for applied work.

The discussion so far has focussed on bias and precision of the different estimators for β. In ordered logit models, the main interest often is not in the β parameters *per-se* (because the latent model may not have a useful interpretation) but rather derived statistics, such as ratios of coefficients or average marginal effects. Ratios can be interesting, because they determine the compensating change of one variable required to offset a change in another one, such that response probabilities remain unchanged. Ratios therefore quantify trade-offs between variables and, if one variable is income or price, monetary compensation. The simulation results for the endogenous cutoff estimators show that the ratio of the means of $\hat{\beta}_1$ and $\hat{\beta}_2$ is far off the true value, and very similar biases are observed if we compute the average of the ratios.

Average marginal probability effects can be computed for an arbitrary value of α_i, such as zero, or at a specific value of the linear index such as the one resulting in the sample probabilities. The average marginal effect of the l-th regressor on the probability of outcome k has the form

$$AME_l^k = (NT)^{-1} \sum_{i,t} - \left[\Lambda_{ik+1}(1 - \Lambda_{ik+1}) - \Lambda_{ik}(1 - \Lambda_{ik})\right] \beta_l,$$

where $\Lambda_{ik} = \Lambda(\tau_{ik} - x_{it}'\beta - \alpha_i)$. Since the AME is proportional to β_l, the relative bias is equal to the relative bias in β_l. For instance, the true average effect of the binary regressor on the highest outcome is a 24.4 percentage point increase in the baseline DGP. Using the

FF estimate, the effect is underestimated by 14 percent.

2.4 The effect of unemployment on life satisfaction

In an early, seminal contribution to the subjective well-being literature, Clark and Oswald (1994) found a strong negative association between a measure of individual psychological well-being (a mental distress score) and current unemployment, using data from the 1991 cross-section of the British Household Panel Survey and adjusting for selected potential socio-economic confounders in ordered probit regressions. A number of subsequent studies confirmed the basic finding and showed that it was robust to modifications of the dependent variable (e.g. self-stated happiness and life satisfaction) or country (see, for instance, Helliwell, 2003, and Blanchflower and Oswald, 2004).

Importantly, the negative association between well-being and unemployment was also shown to be robust to the inclusion of individual-specific fixed effects, countering the objection that the unemployed are inherently less satisfied and that the negative association should therefore not be interpreted as causal (Winkelmann and Winkelmann, 1998). However, such a fixed-effects analysis does not identify the causal effect either if individual and period-specific shocks, for instance a deterioration in an unobserved dimension of health, correlate with both life satisfaction and entrance into unemployment. While one would require an instrument in such a case, in order to isolate the exogenous variation in unemployment, such an estimation strategy has to the best of our knowledge not been pursued in this context before.

Our empirical contribution is therefore twofold. First, we improve on the existing literature by using consistent and efficient fixed effects ordered logit estimators to analyze the relationship between life satisfaction and unemployment. For this analysis, we employ the dataset used in Winkelmann and Winkelmann (1998). The sample was drawn from the first six (1984-1989) waves of the German Socio-economic Panel (GSOEP) and includes observations on men aged 20-64 years who are observed for at least two waves. The outcome variable is satisfaction with life which is measured by the question *"How satisfied are you at present with your life as a whole?"*. The answer has 11 ordered categories ranging from 0, "completely dissatisfied", to 10, "completely satisfied". We re-estimate the original specification using six methods: Chamberlain with fixed cutoff at 8, MD, BUC, FF, Mean and Median.

In a second step, we implement an instrumental variables estimators, with plant closure as an instrument for unemployment. For this analysis, we use a different sample of the GSOEP since plant closure information was collected annually only from 1991 onwards, with the exception of the two years 1999 and 2000. Two related papers, Kassenboehmer and Haisken-DeNew (2009) and Schmitz (2011), also use plant closure to estimate the effect of unemployment on life satisfaction and self-assessed health, respectively. However, they do not instrument for unemployment but rather directly estimate the effect plant closure unemployment, assuming that it is as good as randomly assigned, i.e., that plant closure during the past year *and* being unemployed at the time of the interview is exogenous. To give an example, in our sample for the period 1991 to 2009, there were 560 instances of workers affected by plant closures, and only 239 of those were also unemployed at the interview date. It is likely that these 239 workers are a non-random, self-selected subgroup, invalidating the exogeneity assumption even if loosing ones job due to plant closure initially satisfies the exclusion restriction of an instrument.

2.4.1 Fixed effects ordered logit results

The estimation results are presented in Table 2.2. The key explanatory variables are two dummy variables indicating current labor market status: *Unemployed* and *out of labor force*, *employed* being the omitted reference category. To allow for, possibly non-linear, habituation effects, the specification includes the variables *duration of unemployment* and *squared duration of unemployment*. Marital status (*married*), health status (*good health*), age (*age* and *age squared*) and logarithmic household income (*log household income*) are added as control variables (see Winkelmann and Winkelmann, 1998, for further detail on data and specification).

Every column of Table 2.2 provides results for a different estimator, with standard errors in parentheses. The first column reproduces the original Chamberlain estimates for a dichotomization at value 8. There are 2,573 individuals with variation in d_{it}^8 in an unbalanced panel of 12,980 observations. This dichotomization entails therefore a substantial loss of information, as the total sample had 20,944 person-year observations on 4,261 individuals, of which 3,958 individuals had some variation in the categorical outcome variable. As to the substantive results, the effect of unemployment is found to be large and statistically significant; there is no effect of unemployment duration on life satisfaction.

Columns (2) and (3) of Table 2.2 show the estimates obtained using the MD and BUC methods, and the final three columns show results for the FF, Mean and Median estimators. The efficiency gains of BUC are substantial. For example, the standard error of the unemployment effect drops by 20 percent from 0.20 to 0.16 relative to that of the Chamberlain estimator. MD estimation reduces the standard errors further, to 0.14 in the case of the unemployment effect. The most striking feature of Table 2.2 as a whole is that the first three columns — which are based on consistent estimators — are remarkably similar, while they differ from the three last columns based on inconsistent ones. The marginal effect of unemployment on latent life satisfaction is about -1 when using Chamberlain, MD or BUC but it ranges only from -0.84 to -0.66 when using FF, Mean or Median estimators. A similar attenuation bias is observed for the effects of non-participation, marital status and age.

There is only one noteworthy discrepancy among the consistent estimators. It relates to the coefficient of *out of labor force*, which is -0.24 and insignificant for Chamberlain while being around -0.45 and significant for MD and BUC. A potential explanation for this result is that most changes in *out of labor force* occur at levels of satisfaction lower than the cutoff of 8 used by the Chamberlain estimator, so that this information is not used for estimation. MD and BUC, on the other hand, retain all 3,958 persons displaying some time variation in life satisfaction.

2.4.2 Plant closure, unemployment and life satisfaction

With correlated individual and period-specific shocks, the fixed effects ordered logit estimators of the previous section fail to identify the causal effect of unemployment on happiness. While a potential instrument, plant closure, is available, estimation requires non-standard methods due to the non-linearity of the model. For example, it is not possible to replace the binary unemployment indicator by its predicted value from a first-stage regression. Due to the similarities between ordered and binary response variables, we develop our IV-estimator based on the strand of literature which deals with the case where both endogenous regressor and outcome variable are binary. Dong and Lewbel (2012) discuss different IV-estimators for this case, stressing that all methods have their drawbacks. Maximum likelihood requires a full specification of the joint distribution of the error terms in the selection and outcome equation. Control function approaches require a continuous endogenous variable. The linear probability model is not compatible with the ordered scale of the outcome variable. We therefore use the special regressor approach introduced by Lewbel (2000). Honoré and Lewbel (2002) extend the approach to panel data, and Dong and Lewbel (2012) derive a simple implementable estimator.

In addition to exogenous variation of the instrument, the special regressor approach requires a variable which (i) affects the distribution of d, and (ii) is independent of, and additively separable from, the error term of the linear index model. This variable, denoted by v_{it}, should also have a large support, although this last condition can sometimes be relaxed (see Magnac and Maurin 2007, 2008). Thus if one can write the model as

$$d_{it} = \mathbb{1}(v_{it}\gamma + x'_{it}\beta + \alpha_i + \epsilon_{it} > 0), \tag{2.19}$$

with v_{it} independent of $\alpha_i + \epsilon_{it}$ and $\gamma \neq 0$, the special regressor approach is applicable. Since the parameters of the linear index model are only identified up to scale, the effect of the special regressor is usually normalized to 1. The special regressor approach works by first transforming the nonlinear model into a linear one and then using a linear regression technique to estimate the parameters. Let $\tilde{d}$ denote the linearized dependent variable,

$$\tilde{d} = [d_{it} - \mathbb{1}(v_{it})]/f_t(v_{it}|x_{it}, z_{it}), \tag{2.20}$$

where z_{it} denotes the instrument and $f_t(v_{it}|x_{it}, z_{it})$ the conditional density function of v_{it} which can vary over time. Honoré and Lewbel (2002) show that the expectation of $\tilde{d}$ is linear in x as long as the special regressor is independent of both components of the error term:

$$\mathrm{E}(\tilde{d}_{it}|x_{it}, z_it) = x'_{it}\beta + \mathrm{E}(\alpha_i + \epsilon_{it}|x_{it}, z_{it}). \tag{2.21}$$

If these conditions hold, the linear fixed effects estimator based on $\tilde{d}_{it}$ therefore consistently estimates β as long as ϵ_{it} is orthogonal to the x treated as exogenous and the instrument z.

In our application, the ordered outcome variable “life satisfaction” is first dichotomized into $K-1$ multiple binary variables d^k. $\tilde{d}^k$ is obtained as follows: We regress v_{it} on x_{it} and z_{it}, taking the density function of the residuals to transform d_{it} into $\tilde{d}_{it}$. A separate density function is estimated for each period and the resulting $\tilde{d}$-distribution is trimmed by dropping the smallest and largest percent of the values. Finally, a fixed effects regression instruments employment status with “plant closure” and uses a fixed effect per individual and dichotomization. Our special regressor is the number of interviews with the same interviewer. This choice is a valid special regressor if it has a large support, affects the latent index linearly, and if changing the interviewer is unrelated to employment status. Kassenboehmer and Haisken-DeNew (2012) find that years in panel – a variable closely related to our special regressor – has a strong negative effect on happiness in a linear model with fixed effects. The strong negative effect together with the sizable range of the variable from 1 to 19 ensures the large support condition.

Table 2.3 shows the estimation results. The first four columns use plant closure unemployment directly as a regressor, as in Kassenboehmer and Haisken-DeNew (2009) and Schmitz (2011), while the fifth and final column shows the IV estimates. Sample and specification follow closely these two earlier papers. We use data on men aged 20 to 64 living in West Germany extracted from the GSOEP for the period 1991 to 2009. Two years, 1999 and 2000, had to be excluded from the analysis, since the plant closure question was not asked in these two years. This leaves 17 years, with a total of 82,395 person-year observations. In contrast to the previous section, unemployment duration is dropped, and bi-annual time fixed effects are included.

In contrast to the earlier papers, who used the inconsistent mean cutoff estimator, we report results for the BUC and MD estimators. The first two columns largely corroborate the findings of the earlier section and thus Winkelmann and Winkelmann (1998), using more recent data. The negative point estimate for the effect of unemployment on life satisfaction is even slightly larger. Columns (3) and (4) show the results for a specification with an unemployment main effect and a plant closure – unemployment interaction. The main effect thus gives the effect for those entering unemployment for reasons unrelated to plant closure (such as individual dismissal or quit). The point estimate is virtually unchanged. The plant closure interaction is negative and adds an additional effect of about a quarter.

Importantly, the estimate of the special regressor approach with plant closure as instrument, shown in column (5), confirms the earlier results. To simplify the comparison between BUC, MD and IV approach, we normalized the effect of *out of labor force* in the latter to -0.41. The IV point estimate for unemployment is then -0.99, again a negative effect of similar magnitude. In summary, there are two main conclusions emanating from Table 3. First, there is some evidence that the effect of unemployment is heterogeneous, depending on the reason of entry into unemployment. Second, the similarity of unemployment effects in (1) and (2), the main effects of unemployment in (3) and (4), and the IV estimate in (5) suggest that contemporaneous endogeneity of unemployed is unlikely to be a first-order problem, and that the fixed effects ordered logit results provide reasonable estimates for the average effect of unemployment on life satisfaction.

2.5 Conclusions

The ordered logit model has a number of desirable features that make it the first choice in regression analyses of discrete, ordinally measured variables, as they arise in the elicitation of life- and job satisfaction or self-assessed health. It has a parsimonious yet flexible parametrization that exploits the ordering information while allowing inferences to be made on the entire distribution of outcomes. However, applications of the ordered logit model to panel data with fixed effects have been hampered so far by the lack of a unified discussion of a number of possible estimators and their respective advantages and shortcomings.

We show in this paper that two of the existing approaches used in the prior literature cannot be recommended because they are either inefficient or inconsistent. The bias can be substantial, as shown both in Monte Carlo simulations and in the application to the effect of unemployment on life satisfaction. On the other hand, we derive the consistent and asymptotically efficient GMM estimator that uses all available information and has the same asymptotic covariance matrix as a minimum distance estimator. We also study a modified estimator that is simple to implement and may be more robust in finite samples, although it is not efficient.

Finally, our investigation into the causal effect of unemployment on life satisfaction exploits recent advances in instrumental variables estimation when both outcome variable and endogenous regressor are discrete. We implement a special regressors approach, using plant closure as instrument for unemployment and the number of interviews with the same interviewer as special regressor. Our results corroborate the result of a large negative effect of unemployment on life satisfaction reported in the previous literature.

References

Andersen, Erling B. (1970): "Asymptotic properties of conditional maximum-likelihood estimators", *Journal of the Royal Statistical Society, Series B (Methodological)*, **32**, 283-301.

Blanchflower, D.G., and Andrew J. Oswald (2004): "Well-being over time in Britain and the United States", *Journal of Public Economics*, **88**, 1359-1386.

Booth, Alison L., and Jan C. van Ours (2008): "Job satisfaction and family happiness: the part-time work puzzle", *Economic Journal*, **118**, F77-F99.

Chamberlain, Gary (1980): "Analysis of covariance with qualitative data", *Review of Economic Studies*, **47**, 225-238.

Clark, Andrew E., and Andrew J. Oswald (1994): "Unhappiness and unemployment", *Economic Journal*, **104**, 648-659.

Das, Marcel, and Arthur van Soest (1999): "A panel data model for subjective information on household income growth", *Journal of Economic Behavior & Organization*, **40**, 409-426.

D'Addio, Anna Cristina, Tor Eriksson, and Paul Frijters (2007): "An analysis of the determinants of job satisfaction when individuals' baseline satisfaction levels may differ", *Applied Economics*, **39**, 2413-2423.

Dong, Yingying, and Arthur Lewbel (2012): "A simple estimator for binary choice models with endogenous regressors", Working Papers 111204, University of California-Irvine, Department of Economics.

Ferrer-i-Carbonell, Ada, and Paul Frijters (2004): "How important is methodology for the estimates of the determinants of happiness?", *Economic Journal*, **114**, 641-659.

Frijters, Paul, John P. Haisken-DeNew, and Michael A. Shields (2004 a): "Investigating the patterns and determinants of life satisfaction in Germany following reunification", *Journal of Human Resources*, **39**, 649-674.

Frijters, Paul, John P. Haisken-DeNew, and Michael A. Shields (2004 b): "Money does matter! Evidence from increasing real income and life satisfaction in East Germany following reunification", *American Economic Review*, **94**, 730-740.

Frijters, Paul, John P. Haisken-DeNew, and Michael A. Shields (2005): "The causal effect of income on health: Evidence from German reunification", *Journal of Health Economics*, **24**, 997-1017.

Greene, William H. (2004): "The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects", *Econometrics Journal*, **7**, 98-119.

Helliwell, John F. (2003): "How's life? Combining individual and national variables to explain subjective well-being", *Economic Modelling*, 20(2), 331-360.

Honoré, Bo, and Arthur Lewbel (2002): "Semiparametric binary choice panel data models without strictly exogenous regressors", *Econometrica*, **70**, 2053-2063.

Jones, Andrew M., and Stephanie Schurer (2011): "How does heterogeneity shape the socioeconomic gradient in health satisfaction?", *Journal of Applied Econometrics*, **26**, 549-714.

Kassenboehmer, Sonja C., and John P. Haisken-DeNew (2009): "You're fired! The causal negative effect of unemployment on life satisfaction", *Economic Journal*, **119**, 448-462.

Kassenboehmer, Sonja C., and John P. Haisken-DeNew (2012): "Heresy or enlightenment? The well-being age U-shape effect is flat", *Economics Letters*, **117**, 235 - 238.

Kitamura, Y. (2006): "Empirical likelihood methods in econometrics: Theory and practice", Cowles Foundation discussion paper nr. 1569.

Lancaster, Tony (2000): "The incidental parameter problem since 1948", *Journal of Econometrics*, **95**, 391-413.

Lewbel, Arthur (2000): "Semiparametric qualitative response model estimation with instrumental variables and unknown heteroscedasticity", *Journal of Econometrics*, **97**, 145-177.

Magnac, T., and E. Maurice (2007): "Identification and information in monotone binary models", *Journal of Econometrics*, **139**, 76-104.

Magnac, T., and E. Maurice (2008): "Partial identification in monotone binary models: discrete regressors and interval data", *Review of Economic Studies*, **75**, 835-864.

Mukherjee, Bhramar, Jaeil Ahn, Ivy Liu, Paul J. Rathouz and Brisa N. Sanchez (2008): "Fitting stratified proportional odds models by amalgamating conditional likelihoods", *Statistics in Medicine*. **27**, 4950-4971.

Schmitz, Hendrik (2011): "Why are the unemployed in worse health? The causal effect of unemployment on health", *Labour Economics*, **18**, 71-78.

Winkelmann, Liliana, and Rainer Winkelmann (1998): "Why are the unemployed so unhappy? Evidence from panel data", *Economica*, **65**, 1-15.

A Asymptotic variance of MD estimator

The stacked Chamberlain estimator $\tilde{\beta}$ can be seen as the result of a joint estimation problem with $(K-1)\cdot L$ parameters. Since the estimator is consistent but the dependence between contributions of the same individual are not incorporated in the maximization problem, this is an instance of quasi-maximum likelihood estimation. The asymptotic variance has therefore the form:

$$\mathrm{AVar}(\tilde{\beta}) = \mathrm{E}(H_i^{CH}(\beta))^{-1}\mathrm{E}(s_i(\beta)s_i(\beta)')\mathrm{E}(H_i^{CH}(\beta))^{-1},$$

where $s_i(\beta)$ is the stacked Chamberlain scores, and $H_i^{CH}(\beta)$ denotes the Hessian. Each likelihood contribution depends only on one set of L parameters. The Hessian of the joint parameter vector is hence block diagonal and $M'H_i^{CH}(\beta) = H_i(\beta)'$, where $H_i(\beta)$ denotes the stacked Chamberlain Hessians, and M the matrix of $K-1$ stacked L-dimensional identity matrices. Therefore, the asymptotic variance of the minimum distance estimator can be written as

$$\begin{aligned}
\mathrm{AVar}(\hat{\beta}^{MD}) &= (M'\mathrm{AVar}(\tilde{\beta})^{-1}M)^{-1} \\
&= (M'\mathrm{E}(H_i^{CH}(\beta))\mathrm{E}(s_i(\beta)s_i(\beta)')^{-1}\mathrm{E}(H_i^{CH}(\beta))M)^{-1} \\
&= (\mathrm{E}(H_i(\beta))'\mathrm{E}(s_i(\beta)s_i(\beta)')^{-1}\mathrm{E}(H_i(\beta)))^{-1}.
\end{aligned}$$

Table 2.1: Monte Carlo simulation results (1000 replications)

	N=500, T=3, K=5				*N=1,000, T=3, K=5*				*N=500, T=6, K=5*				*N=500, T=3, K=10*			
	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_1$		$\hat{\beta}_2$	
Chamberlain estimators																
$y \geq 2$	1.00	(0.16)	1.01	(0.26)	1.01	(0.11)	1.01	(0.18)	1.00	(0.10)	1.00	(0.16)	1.01	(0.18)	1.02	(0.36)
$y \geq 3$	1.00	(0.15)	1.02	(0.21)	1.01	(0.11)	1.01	(0.14)	1.00	(0.09)	1.00	(0.13)	1.00	(0.16)	1.01	(0.26)
$y \geq 4$	1.00	(0.15)	1.01	(0.21)	1.00	(0.11)	1.01	(0.14)	1.00	(0.09)	1.00	(0.13)	1.00	(0.16)	1.01	(0.23)
$y \geq 5$	1.00	(0.16)	1.03	(0.26)	1.00	(0.11)	1.01	(0.18)	1.00	(0.10)	1.01	(0.16)	1.00	(0.15)	1.02	(0.21)
Estimators with endogenous cutoffs																
FF	0.92	(0.13)	0.86	(0.18)	0.92	(0.10)	0.86	(0.12)	0.96	(0.08)	0.92	(0.11)	0.88	(0.13)	0.80	(0.17)
Median	0.92	(0.13)	0.88	(0.18)	0.93	(0.10)	0.87	(0.12)	0.96	(0.09)	0.92	(0.11)	0.89	(0.13)	0.83	(0.17)
Mean	0.95	(0.13)	0.92	(0.18)	0.95	(0.10)	0.91	(0.12)	0.97	(0.09)	0.95	(0.12)	0.94	(0.13)	0.90	(0.18)
Estimators which combine all information																
MD	0.98	(0.14)	1.00	(0.18)	1.00	(0.10)	1.00	(0.12)	0.99	(0.09)	1.00	(0.11)	0.95	(0.13)	0.98	(0.17)
BUC	1.00	(0.14)	1.01	(0.18)	1.00	(0.10)	1.01	(0.12)	1.00	(0.09)	1.01	(0.11)	1.00	(0.14)	1.01	(0.17)
GMM	1.02	(0.14)	1.04	(0.19)	1.01	(0.10)	1.02	(0.12)	1.00	(0.09)	1.00	(0.11)	1.03	(0.15)	1.07	(0.19)
EL	1.00	(0.14)	1.01	(0.18)	1.01	(0.10)	1.01	(0.12)	1.00	(0.09)	1.00	(0.11)	1.00	(0.14)	1.01	(0.18)

Notes: $\beta_1 = \beta_2 = 1$. Columns contain the mean of the estimated coefficients over all replications where the estimator has converged, standard deviation of the estimated coefficients in paranthesis. Number of cases where a estimator did not converge: *N=500, T=3, K=5*: GMM: 7, EL:1; *N=1'000, T=3, K=5*: GMM: 3; *N=500, T=6, K=5*: GMM: 9; *N=500, T=3, K=10*: GMM: 8.

Table 2.2: Fixed Effects Ordered Logit Estimates of Life Satisfaction

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.: *Life Satisfaction*	$y \geq 8$	MD	BUC	FF	Mean	Median
Unemployed	-0.96**	-0.98**	-1.03**	-0.77**	-0.84**	-0.66**
	(0.20)	(0.14)	(0.16)	(0.15)	(0.15)	(0.15)
Out of labor force	-0.24*	-0.42**	-0.45**	-0.25**	-0.25**	-0.25**
	(0.12)	(0.09)	(0.11)	(0.09)	(0.10)	(0.09)
Duration of unemployment	-0.01	-0.01	-0.02*	-0.02*	-0.01	-0.01
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Squared duration of unemp.	0.60	2.44	2.75	3.18	2.17	2.12
$\times 10^{-4}$	(2.79)	(1.56)	(2.30)	(1.87)	(1.88)	(1.86)
Married	0.67**	0.52**	0.56**	0.37**	0.39**	0.37**
	(0.12)	(0.09)	(0.11)	(0.09)	(0.09)	(0.09)
Good health	0.34**	0.33**	0.36**	0.24**	0.29**	0.24**
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Age	-0.12**	-0.12**	-0.12**	-0.12**	-0.11**	-0.12**
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Age squared $\times 10^{-2}$	-0.84	-2.46	-1.15	-1.30	-2.91	-1.58
	(4.27)	(3.24)	(3.82)	(3.36)	(3.38)	(3.35)
Log household income	0.13**	0.12**	0.13**	0.10**	0.10**	0.10**
	(0.06)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)
Log likelihood	-4,996	—	-21,802	-8,003	-7,911	-8,054
Observations	12,980	59,535	59,535	19,053	19,071	19,071
Individuals	2,573	3,958	3,958	3,949	3,958	3,958

Notes: Data source GSOEP, waves 1984-1989; cluster robust standard errors in parentheses; */** indicates statistical significance at the 10%/1% level. "Observations" denotes the number of person-years in estimation sample; "Individuals" denotes number of unique persons in estimation sample.

Table 2.3: Plant closure, unemployment and life satisfaction

	(1)	(2)	(3)	(4)	(5)
Dep. var.: *Life Satisfaction*	BUC	MD	BUC	MD	IV
Unemployed	-1.12**	-1.05**	-1.10**	-1.03**	-0.99*
	(0.05)	(0.05)	(0.05)	(0.05)	(0.59)
Unemployed $\times$ *Plant Closure*			-0.27*	-0.28*	
			(0.16)	(0.16)	
Out of labor force	-0.41**	-0.35**	-0.41**	-0.35**	-0.41**
	(0.05)	(0.04)	(0.05)	(0.04)	(0.14)
Married	0.32**	0.36**	0.32**	0.36**	0.27**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.08)
Good health	0.31**	0.30**	0.31**	0.30**	0.13**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Age	-0.17**	-0.19**	-0.17**	-0.19**	-0.19**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Age squared $\times 10^{-2}$	0.06**	0.07**	0.06**	0.06**	0.11**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Log household income	0.26**	0.26**	0.26**	0.26**	0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)
Log likelihood	-84,024		-84,019		
Observations	231,657		231,657		607,059
Individuals	9,720	9,722	9,720	9,722	11,979

Notes: GSOEP, waves 1991-2009 without the years 1999 and 2000. All regressions include bi-annual time fixed effects. BUC and MD refer to the ordered logit fixed effects Blow-up-and-Cluster and Minimum Distance estimators, respectively. The IV-results are computed using *plant closure* as instrument for *unemployed* and number of interviews with the same interviewer as special regressor. The IV-coefficients are normalized by setting the effect of *out of labor force* to -0.41. Otherwise the notes of Table 2.2 apply.

Chapter 3

Identification and estimation of thresholds in the fixed effects ordered logit model

This chapter has been published in *Economics Letters*, 115, pp. 416-418.

Acknowledgements: I thank Kevin Staub and Rainer Winkelmann for very valuable comments.

3.1 Introduction

There exist a number of proposals in the literature on how to estimate a panel-ordered logit model with individual fixed effects – Das and van Soest (1999), Ferrer-i-Carbonell and Frijters (2004) and most recently Baetschmann, Staub and Winkelmann (2011). A drawback of all these estimators is that they do not identify the threshold parameters. This paper proposes a new estimating procedure which allows estimating these thresholds. Knowing the thresholds has three advantages: First, the thresholds are helpful for interpreting the regression coefficients; second, they make it possible to obtain statements about the effect of a changing x on the observed ordered variable and not only on the latent variable; and third, comparing the differences between the thresholds can be interesting in itself. The new procedure can be easily implemented using existing software for conditional maximum likelihood (CML) logit estimation with cluster corrected standard errors.

The paper proceeds as follows. Section 3.2 presents the fixed effects ordered logit model and discusses the new estimation procedure. In section 3.3 the new estimator is applied to data from the German Socioeconomic Panel.

3.2 Econometric Methods

3.2.1 The FE ordered logit model

The fixed effects ordered logit model relates the latent variable y_{it}^* for individual i at time t to a linear index of observable characteristics x_{it} and unobservable characteristics α_i and ε_{it}:

$$y_{it}^* = x_{it}'\beta + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N \quad t = 1, \dots, T. \tag{3.1}$$

The time-invariant part of the unobservables (α_i), called fixed effect, can be statistically dependent of x_{it}.

The latent variable is tied to the (observed) ordered variable y_{it} by the observation rule:

$$y_{it} = k \quad \text{if} \quad \tau_k < y^*_{it} \leq \tau_{k+1}, \quad k = 1, \ldots, K, \tag{3.2}$$

where the thresholds τ are assumed to be strictly increasing ($\tau_k < \tau_{k+1} \; \forall \, k$) and $\tau_1 = -\infty$, $\tau_{K+1} = \infty$.

The specification of the fixed effects ordered logit model is completed by assuming that the ε_{it} are conditionally independent and identically standard logistically distributed. I.e., if $F(\cdot)$ denotes the cdf

$$F(\varepsilon_{it}|x, \alpha) = \frac{\exp(\varepsilon_{it})}{1 + \exp(\varepsilon_{it})} \equiv \Lambda(\varepsilon_{it}). \tag{3.3}$$

Hence, the probability of observing an outcome equal to k for individual i at time t using (3.1), (3.2) and (3.3) can be written as

$$\Pr(y_{it} = k|x_{it}, x'_{it}\beta) = \Lambda(\tau_{k+1} - x'_{it}\beta - \alpha_i) - \Lambda(\tau_k - x'_{it}\beta - \alpha_i), \tag{3.4}$$

whereas the probability of an outcome greater or equal to k is

$$\Pr(y_{it} \geq k|x_{it}, x'_{it}\beta) = \Lambda(x'_{it}\beta + \alpha_i - \tau_k). \tag{3.5}$$

Equation (3.4) and (3.5) show that the location of the τ's and α's cannot be distinguished. Thus the constant and the second threshold (τ_2) are normalized to zero.

The problem with maximum likelihood estimation based on (3.4) is that the expression depends on the individual fixed effect α_i. Including individual dummies in the estimation procedure to account for fixed effects is not a solution due to the "incidental parameter problem" – e.g. Chamberlain (1980).

3.2.2 Illustration of the estimation procedure

The binary logit model is one of the few nonlinear models, where it is known how to deal with fixed effects. For this model, Chamberlain (1980) proposed to condition the likelihood on the number of one's in individual's record to get rid of the individual fixed effects. Chamberlain's method can be applied to the ordered logit model as well. The procedure is as follows: First, the ordered dependent variable y is dichotomized to a binary one. The binary variable is denoted by d and the cutoff by k: $d = 1(y \geq k)$, where $1()$ is the indicator function. Second, Chamberlain's estimation procedure is applied to d.

To illustrate this procedure consider an individual which is observed two times, where the first observation equals 3 and the second equals 1: $y_1 = 3$, $y_2 = 1$. We assume that the variable y can take the values 1, 2 and 3, thus k can be either 2 or 3. In this example both choices result in the same binary dependent variable, $d_1 = 1$ and $d_2 = 0$. The following conditional probability results:

$$\begin{aligned}
&\Pr\big[d_1 = 1 \cap d_2 = 0 \big| (d_1 = 1 \cap d_2 = 0) \cup (d_1 = 0 \cap d_2 = 1)\big] \\
&\quad = \Pr\big[y_1 \geq k \cap y_2 < k \big| (y_1 \geq k \cap y_2 < k) \cup (y_1 < k \cap y_2 \geq k)\big] \\
&\quad = \frac{\frac{\exp(x_1'\beta+\alpha-\tau_k)}{1+\exp(x_1'\beta+\alpha-\tau_k)} \frac{1}{1+\exp(x_2'\beta+\alpha-\tau_k)}}{\frac{\exp(x_1'\beta+\alpha-\tau_k)}{1+\exp(x_1'\beta+\alpha-\tau_k)} \frac{1}{1+\exp(x_2'\beta+\alpha-\tau_k)} + \frac{1}{1+\exp(x_1'\beta+\alpha-\tau_k)} \frac{\exp(x_2'\beta+\alpha-\tau_k)}{1+\exp(x_2'\beta+\alpha-\tau_k)}} \\
&\quad = \frac{\exp(x_1'\beta + \alpha - \tau_k)}{\exp(x_1'\beta + \alpha - \tau_k) + \exp(x_2'\beta + \alpha - \tau_k)} = \frac{\exp(x_1'\beta)}{\exp(x_1'\beta) + \exp(x_2'\beta)}. \qquad (3.6)
\end{aligned}$$

The last expression in (3.6) is independent of α. Thus β can be estimated by conditional maximum likelihood.

The problem with this procedure is that the τ's disappear from the probability expression as well and are therefore not identified. The reason is that the same cutoff is used for all observations of an individual, so there is no "cutoff-variation" within a conditional likelihood contribution. By contrast, if the observations of an individual are dichotomized at different cutoff points and the probability expression is applied accordingly, the thresholds are identified. Consider again the above example, but suppose now that the first observation is dichotomized at 2 and the second at 3. The probability that the the first dichotomization is one, given that either the first or the second is one (but not both) is now:

$$\Pr\big[y_1 \geq 2 \cap y_2 < 3\big|(y_1 \geq 2 \cap y_2 < 3) \cup (y_1 < 2 \cap y_2 \geq 3)\big]$$
$$= \frac{\frac{\exp(x_1'\beta+\alpha-\tau_2)}{1+\exp(x_1'\beta+\alpha-\tau_2)}\frac{1}{1+\exp(x_2'\beta+\alpha-\tau_3)}}{\frac{\exp(x_1'\beta+\alpha-\tau_2)}{1+\exp(x_1'\beta+\alpha-\tau_2)}\frac{1}{1+\exp(x_2'\beta+\alpha-\tau_3)}+\frac{1}{1+\exp(x_1'\beta+\alpha-\tau_2)}\frac{\exp(x_2'\beta+\alpha-\tau_3)}{1+\exp(x_2'\beta+\alpha-\tau_3)}}$$
$$= \frac{\exp(x_1'\beta+\alpha-\tau_2)}{\exp(x_1'\beta+\alpha-\tau_2)+\exp(x_2'\beta+\alpha-\tau_3)} = \frac{\exp(x_1'\beta-\tau_2)}{\exp(x_1'\beta-\tau_2)+\exp(x_2'\beta-\tau_3)}. \tag{3.7}$$

This expression is independent of α but depends on β and the τ's. Hence there is no incidental parameter problem, and β and τ can be estimated by maximum likelihood. The method can easily be generalized to situations with more than two time periods and more than two possible cutoff points.

3.2.3 Practical implementation – choosing the cutoff points

The question arises, which combinations of observation specific cutoff points to include in the estimation procedure. One possibility is to include all feasible combinations. Cluster standard errors can be used to account for the dependence between the conditional likelihood contributions of the same individual (White, 1982). The same idea of including more than one "clone" of an individual combined with cluster standard errors is used by the BUC estimator (Baetschmann, Staub and Winkelmann, 2011) to estimate β in the FE ordered logit model with individual specific thresholds. In the previous example, there are four combinations: Both observations can be dichotomized at two cutoff points and all four combinations are possible. Among those, only the two combinations with different cutoff points are informative for estimating the τ's.

The inclusion of all possible cutoff point combinations in the estimation procedure is only feasible if the number of time periods (T) and the number of categories (K) of the dependent variable are small, because the number of possible combinations is $(K-1)^T$. For example if T and K are equal to 10, there exist more than three billion possible copies of each individual. Often the researcher is more interested in estimating β than τ. On this account, I propose to include all clones with no variation in the cutoff to estimate β precisely and fill up the rest of the dataset with a limited number of clones with random variation in the cutoff points. (Stata code is available from the author upon request.)

3.3 Illustration

To illustrate the estimation procedure, the new fixed effects ordered logit estimator is applied to the model and dataset of Winkelmann and Winkelmann (1998). The dataset consists of a sample from the German Socioeconomic Panel going from 1984 to 1989 with 4'261 individuals. The dependent variable is satisfaction with life, which is measured as answer to the question "How satisfied are you at present with your life as a whole?". The answers ranges from 0, "completely dissatisfied", to 10, "completely satisfied". To be consistent with the notation of the theoretical part of this paper, the dependent variable is recoded and ranges now from 1 to 11.

If each individual would be dichotomized in all possible ways, the resulting dataset would consist of more than four billion entries. Hence I decided to include all clones with a constant cutoff to estimate β precisely, plus ten clones of each individual, whose observations are dichotomized at observation specific random cutoff points. Compared to other proposed estimators, individuals without variation in the ordered dependent variable are not automatically excluded from the estimation procedure. The reason is that variation in the ordered dependent variable (y) is not a precondition for variation in the dichotomized dependent variable (d) if varying cutoffs within a conditional likelihood contribution are allowed.

The columns with heading "$y \geq 8$" of Table 3.1 show the estimates of the Chamberlain "estimator" if the ordered variable is dichotomized at 8. These estimates are also reported in Winkelmann and Winkelmann (1998). The results of the new estimation procedure are very similar and are listed in the columns with heading "BUC$_\tau$". The standard errors of the new estimator are slightly smaller. The essential advantage of the new estimation procedure is that estimates for the thresholds are available. The differences between them ranges from 0.66 between τ_3 and τ_4 to 1.69 between τ_9 and τ_{10}. Roughly speaking, the differences between the thresholds increase with the threshold number. This means that the effect of an increasing latent index – for example, by +1.10 when being employed rather than unemployed – on the ordered life satisfaction variable is largest for people with a low life satisfaction level.

References

Baetschmann, Gregori, Kevin E. Staub, and Rainer Winkelmann (2011): "Consistent estimation of the fixed effects ordered logit model", IZA Discussion Paper No. 5443.

Das, Marcel, and Arthur van Soest (1999): "A panel data model for subjective information on household income growth", *Journal of Economic Behavior & Organization*, **40**, 409-426.

Chamberlain, Gary (1980): "Analysis of covariance with qualitative data", *Review of Economic Studies*, **47**, 225-238.

Ferrer-i-Carbonell, Ada, and Paul Frijters (2004): "How important is methodology for the estimates of the determinants of happiness?", *Economic Journal*, **114**, 641-659.

White, Halbert (1982): "Maximum likelihood estimation of misspecified models", *Econometrica*, **50**, 1-25.

Winkelmann, Liliana, and Rainer Winkelmann (1998): "Why are the unemployed so unhappy? Evidence from panel data", *Economica*, **65**, 1-15.

Table 3.1: Fixed Effects Ordered Logit Estimates of Happiness

Dep. var.: *Life Satisfaction*	$y \geq 8$		BUC$_\tau$	
Unemployed	-0.96**	(0.22)	-1.10**	(0.18)
Out of labor force	-0.24	(0.13)	-0.48**	(0.12)
Duration of unemployment	-0.01	(0.02)	-0.02	(0.02)
Squared duration of unemp. $\times 10^{-4}$	0.60	(3.54)	2.45	(2.54)
Married	0.67**	(0.14)	0.59**	(0.12)
Good health	0.34**	(0.06)	0.35**	(0.05)
Age	-0.12**	(0.04)	-0.11**	(0.03)
Squared age $\times 10^{-2}$	-0.84	(4.56)	-1.07	(3.95)
Log. household income	0.13*	(0.06)	0.13*	(0.05)
τ_3			0.69	(0.15)
τ_4			1.35	(0.15)
τ_5			2.17	(0.15)
τ_6			2.90	(0.15)
τ_7			4.31	(0.16)
τ_8			5.06	(0.16)
τ_9			6.28	(0.16)
τ_{10}			7.97	(0.17)
τ_{11}			9.17	(0.17)
Observations	12'980		204'574	
Individuals	2'573		4'204	

Notes: Data Source GSOEP, waves 1984-1989. * (**) statistical significance at 5% (1%) level; Cluster robust standard errors in parentheses.

Chapter 4

Heterogeneity in the relationship between happiness and age: Evidence from the German Socio-Economic Panel

A version of this chapter has been published in *German Economic Review*, forthcoming.

Acknowledgements: I would like to thank Rainer Winkelmann for many helpful suggestions which improved this article.

4.1 Introduction

This paper contributes to the recent literature on the evolution of individual satisfaction over the life cycle.[1] The most prominent hypothesis is that of a U-shape relation between age and happiness. Detailed studies of the relationship, especially for Germany, have confirmed the U-shape over a long range of the life course, but have found another downturn at the end of life (Wunder et al. (2009), Van Landeghem (2009), Fischer (2009), Gwozdz and Sousa-Poza (2009)). Knowing how life satisfaction evolves helps to answer questions like "what is the probability that well-being decreases in the next ten years for a currently 40 year old woman?" or "how happy will I be in twenty years?". Further, it can help to optimize saving decisions. For instance, taking into account the U-shape could help people avoid oversaving for old age.

A key shortcoming of the previous literature is the neglect of heterogeneity in the relationship between age and life satisfaction. With few exceptions (Mroczek and Spiro (2005), Schilling (2006)), the conducted studies have only looked at the central tendency of well-being over the life course at the aggregate level. But seeing a U-shape for the average does not mean that such a relationship is representative for the individual. It is possible that a U-shape results from averaging over individual life cycle paths, which are themselves not U-shaped. Moreover, it is possible that age influences the whole distribution of life satisfaction and not only the location.

Using the longest running panel household survey with continuous information on life satisfaction so far, the German Socio-Economic Panel (GSOEP), it becomes possible to trace individual satisfaction levels for up to 26 years, and hence, in principle, to estimate the relationship between age and life satisfaction at the individual level using time series methods. A further advantage of the GSOEP data is that they include information on the entire adult population (those age 20 or above), including the very old. For example seven

[1]The terms life satisfaction, happiness and well-being are used interchangeable in this paper. For a summary of the literature, see for example Blanchflower and Oswald (2007).

per cent of the sample (more than 20'000 observations) are older than 70 years. This broad coverage is important for testing the hypothesis of a second turning point, i.e. a decrease in life satisfaction at high age.

The key contributions of the paper are as follows: First, I replicate the findings in the literature regarding the relationship between age and average life satisfaction in a general semi-parametric model. Second, I provide an extended analysis of heterogeneity in the life course of satisfaction using evidence from four types of models: dispersion as dependent variable; analysis by subgroups; latent class analysis; and individual level regressions.

The findings of the paper are compatible with a U-shape over most of a person's adult life time. However, a more detailed analysis reveals a mild downward trend up to around 55, followed by a distinct increase. After the age of 70, the curve is clearly falling. The study of the distribution of well-being shows a mixed picture. The fraction of people with a very high satisfaction level is falling over the life course. To a lesser extent, this is also true for the fraction of people with a low satisfaction level. Combined, this results in a decreasing dispersion in life satisfaction between people over the life course.

Whereas men and women show a very similar development over the life course, education groups differ strongly. People with low education seem to suffer from a steady decline in life satisfaction, while well educated people become happier. However, the hump shape after 55 can be found in all education groups. The result of the finite mixture model confirms the hypothesis that heterogeneity between people can be primarily found in the trend over the life cycle, whereas less heterogeneity exists in the hump shape after 55. Another important finding is that the length a person spent in the panel strongly affects the response. This duration effect, the high variance at the individual level, and the rough measurement of life satisfaction are probably the reasons why investigating the relationship between age and life satisfaction at the individual level provides no clear insights.

4.2 Modeling life satisfaction over time

Any study attempting to identify and estimate the relationship between age and life satisfaction needs to take a stance on a number of issues. First, what variables to condition on; second, how to deal with the fundamental identification problem between the effects of age, cohort and time, and whether or not to include individual fixed effects; third, to define the relevant level of aggregation; and fourth, what assumptions to make regarding the econometric model, for a given set of regressors: parametric versus semi-parametric, and linear regression versus non-linear ordered probit or ordered logit models. The following sub-sections provide a discussion of each of these four points, their treatment in the existing literature as well as the position adopted in the present paper.

4.2.1 Conditional vs. unconditional effects

This paper focuses on the question of how life satisfaction has evolved over the life course in Germany in recent years, because I think that results from such an inquiry can be extrapolated and help predicting the evolution of life satisfaction in the near future in Germany and other advanced industrial countries. To get an answer to the question of how life satisfaction has evolved, one essentially needs to follow different people and record their well-being levels. This is the unconditional approach.

In the conditional approach, the researcher is trying to hold some individual level characteristics – like income, health or marital status – constant in order to get a "ceteris paribus" interpretation. But these variables are potential channels through which age affects life satisfaction. Holding individual level variables constant can therefore be highly misleading. Comparing a 50 year old man with three children and a monthly income of \$20,000 with a man with the same characteristics but only age 20, hardly helps to identify the effect of age on life satisfaction. The focus on the unconditional age effect is in line with the view expressed by Glenn (2009), Easterlin (2006), and Easterlin and Sawangfa

(2007), and in contrast to the approach of Blanchflower and Oswald (2008). Of course, if it is the goal of the analysis to identify the channels through which age influences life satisfaction, it is meaningful to include individual level covariates. This paper concentrates on the evolution of well-being over the life course *per-se*, and not on the causal channels.

Nevertheless, there are some variables for which one should control in the econometric model. Year of birth is correlated with age and has probably also an effect on life satisfaction, but is surely not a channel through which age influences happiness. Thus the econometric model should control for cohort effects (see e.g. Blanchflower and Oswald (2008) for a discussion). There is also recent evidence that "panel learning" can have a substantial effect on the response behavior by persons. Panel learning means that people change their responses over time just because they have participated repeatedly in the survey, i.e., even if the underlying feature one wants to measure is unchanged. In the context of life satisfaction, Kassenboehmer and Haisken-DeNew (2010) found a negative effect of time spent in the panel. They conjecture that confidence in the interviewer may rise with each additional interview, which leads to more honest (in this case lower) answers to the life satisfaction question. Interestingly, this panel duration effect has been ignored by much of the previous literature, including the studies by Wunder et al. (2009) and Ree and Alessie (2011), putting a serious question mark behind the findings of these papers.

A further controversial question is whether or not one should control for time effects. The time effect is often split into a shock (for example business cycle) and a trend (e.g. long run economic growth). For example, if the observation period falls together with an economic recession, it looks as if people become unhappier as they get older. However, if the observation period is long enough and different cohorts are tracked, there is no reason why the shocks should be correlated with age. Regarding the long-term trend, it seems pointless to compute age-profiles that exclude the long-term trend, as time and age move in unisono. Arguably, therefore, one should not condition on the trend, but rather focus on the combined effect of age and time in order to predict future age-profiles.

4.2.2 A fundamental identification problem

In an additive model with age, cohort and time, it is not possible to disentangle the linear effect of these three variables, whereas the deviation from the linear effects of each of the three variables is identified. This was formally shown in McKenzie (2006) in a general non-parametric framework. The essence of the argument can be demonstrated in a simple model with linear and quadratic terms:

$$y = \beta_0 + \beta_a age + \gamma_a age^2 + \beta_c cohort + \gamma_c cohort^2 + \beta_t time + \gamma_t time^2 + \epsilon \tag{4.1}$$

Because $age = cohort + time$, there exists a multicollinearity problem between the three variables. This means that it is not possible to identify β_a, β_c, and β_t separately. If one of the three linear terms is dropped, the regression can be run, but the coefficient on the other two remaining linear terms combine their own effect and the effect of the dropped variable. Clearly, $age^2 \neq cohort^2 + time^2$ (unless one of the terms on the right is zero), and hence there is no problem of multicollinearity here. The same holds true for higher order terms. Thus the linear effect of the three variables cannot be disentangled, whereas deviations from the linear effects are identified.

Ree and Alessie (2011) argue that it is thus not possible to assess the hypothesis of a U-shape but only the hypothesis of a convex relationship between age and happiness. Convexity is a weaker claim than U-shape. In the simple example above where happiness is a linear function of age and age^2, convexity means that $\gamma_a > 0$. A U-shape relationship further requires that a minimum exists ($\beta_a < 0$), and that the minimum lies in the observed age range, thus $-\beta_a/(2\gamma_a) \in [20, 80]$. Convexity is necessary but not sufficient for a U-shape. If β_a is not identified, it is hence only possible to test one requirement of a U-shape, namely the convexity, but not to test for a U-shape itself. Of course, if convexity is rejected, then so is the U-shape.

The argument of Ree and Alessie (2011) is formally correct. However, one has to ask, if this unidentified isolated age effect is really the effect we are interested in. As argued above, this age effect, which is fully disentangled from the time effect, is not interesting and has no clear interpretation. The underlying reason for the lack of a meaningful or causal interpretation is that it is not possible to become older without proceeding in time. In contrast, the age effect combined with the estimated linear time effect is interesting and useful. For example, if a person wants to predict his well-being level in ten years, he is not interested in the isolated effect of age but in the total effect of age and time. For him, it is not meaningful to assume that the social and economic conditions, for which the time variable is a proxy for, will be the same as today. And the best estimate for the effect of these changing conditions in the future is probably the linear time effect in the last years. This effect is estimated by dropping *time* from (4.1), in which case *age* estimates $\beta_a + \beta_c$.

Another question is if one should include individual fixed effects into the econometric model. It can be argued that this does more harm than good in the present case. First, there is no obvious reason (for example a selection problem), why one should include individual fixed effects into the econometric model. If the sampling process and the cohorts are stable over time, the cohort variable will control for systematic correlations between age and the individual fixed effects. Second, including individual fixed effects into the econometric model leads to a high conditional correlation between age and panel duration. The panel duration effect would then be identified only from people who do not take part in the survey in one year but return in the next. There are relatively few such cases. Without individual fixed effects, the main variation in the panel duration results from various refreshment samples, where new subjects of various ages were recruited at different points in time.

4.2.3 Aggregation problem

Phenomena at the aggregate level or mean effects are often crucial for policy recommendations. However, for explaining and understanding a phenomenon in the aggregate, it is important to link them to patterns on the individual level. In the context of the relation between age and happiness, the difficulty is that different mixtures of distinct individual life course paths can lead to the same aggregate pattern. This problem is illustrated in Figure 4.1. In scenario A, the population consists of two types, both with a share of 50%. For both types, the evolution of life satisfaction is U-shaped, but they differ in the level and curvature. In scenario B, the population consists again of two different types with equal shares. In contrast to the first one, none of the two types has a U-shaped pattern. However, the relation between age and life satisfaction in the aggregate, as represented by the solid line, is the same in both scenarios.

It is simply not possible to say something about patterns at the individual level by only looking at the aggregate picture. Thus "midlife crisis", for example, would only be a valid explanation for the U-shape in the aggregate in scenario A. An explanation for the second scenario would be that people differ in their discount rates and can choose between two different life cycle paths. People with a high discount rate choose the path with the higher initial life satisfaction level, while people with a low discount rate choose the path with the higher average score.[2]

The literature so far has focused on the aggregate pattern. To the best of my knowledge, the only studies in the age-happiness literature which give some attention to this problem are Mroczek and Spiro (2005) and Schilling (2006). Another study in the happiness literature touching this problem is the paper of Clark et al. (2005), which tries to capture heterogeneity in the income effect on life satisfaction with a finite mixture model.

4.2.4 Econometric model and estimation

To investigate the relation between age and life satisfaction, an additive model is used throughout the paper. In this section, I describe the basic version of the model that focuses on aggregate patterns, as represented by the conditional expectation. In later sections, the model will be modified appropriately in order to enable the study of heterogeneity.

[2]This argument assumes that satisfaction is period specific, i.e., anticipation of future increases or reductions in satisfaction do not enter present satisfaction.

As discussed previously, the included regressors are age (a), year of birth (c for cohort), year of the interview (t for time) and the time spent in the panel up to the interview (d for duration). The dependent variable is a measure of life satisfaction and is denoted by y. A flexible additive model for the expectation of y can be written as:

$$\mathrm{E}(y|a,c,t,d) = \beta_0 + \sum_{k=20}^{80} \beta_k^a I_k^a + \sum_{k=1904}^{1989} \beta_k^c I_k^c + \sum_{k=1984}^{2009} \beta_k^t I_k^t + \sum_{k=1}^{26} \beta_k^d I_k^d, \tag{4.2}$$

where β_0 denotes the constant and I the indicator function (thus I_k^a, for example, equals 1 if the age variable is equal to value of the indicator k). The model includes a dummy for each category of the four variables, and β stands for the effect of the corresponding dummy.

Two sorts of restrictions have to be imposed to enable estimation:

$$\sum_{k=\min(x)}^{\max(x)} \beta_k^x = 0 \qquad \text{for all } x \in \{a,c,t,d\} \tag{4.3}$$

$$\sum_{k=1984}^{2009} \beta_k^t k = 0. \tag{4.4}$$

Equation (4.3) restricts the total effect of each variable to zero and hence avoids multicollinearity between the dummies. It is functionally equivalent to dropping one dummy of each variable. The second restriction – equation (4.4) – ensures that the linear effect of the time variable is equal to zero and thus avoids multicollinearity between the linear effects of age, cohort, and time (cf. Section 4.2.2).

The identified linear effects can be directly estimated by reformulating the econometric model. Including a separate term for the trend of each variable and using the identity $time = age + cohort$ results into the following model:

$$\begin{aligned}\mathrm{E}(y|a,c,t,d) =& \beta_0 + (\beta^a + \beta^t)a + \sum_{k=20}^{80} \beta_k^a I_k^a + (\beta^c + \beta^t)c + \sum_{k=1904}^{1989} \beta_k^c I_k^c + \\ & \sum_{k=1984}^{2009} \beta_k^t I_k^t + \beta^d d + \sum_{k=1}^{26} \beta_k^d I_k^d,\end{aligned} \tag{4.2'}$$

where β's without subscript denote the trends. To enable estimation, restriction (4.3) and restrictions equivalent to (4.4) on cohort, age and duration in addition to time are imposed.

Because the variable *time* was replaced by *age* plus *cohort*, the variables *age* and *cohort* estimate their own linear effect plus the time trend.

Both formulas represent the same model, which can be estimated by ordinary least squares (OLS). The well-being variable is usually described as ordered, and the median is normally viewed as the right statistic to characterize the location of an ordered variable.[3] The mean, however, has the advantage to be more sensitive to small changes in the distribution. Previous studies – Ree and Alessie (2011), Wunder et al. (2009), and Van Landeghem (2009) – have found a magnitude of less than one point on the eleven point scale in Germany. Thus it is possible that the location of the distribution changes systematically over the life course, whereas the median is constant. For this reason, the mean is used to study the evolution of average life satisfaction. To allow for dependence between repeated observations of one individual, cluster corrected standard errors are reported.

4.3 Results

4.3.1 Data description

The paper uses data from the German Socio Economic Panel (GSOEP). The unique feature of this data set is its long time dimension. At present, it is possible to follow some people for 26 years, from 1984 to 2009. The analysis is conducted with unweighted observations of people, who live in (former) West-Germany and are between 20 and 80 years old. Life satisfaction is ascertained with the question "How satisfied are you with your life, all things considered?" that is always asked at the end of the interview.[4] The response is measured on an eleven point scale ranging from 0 (completely dissatisfied) to 10 (completely satisfied).

[3]Of course, one could also estimate conditional probability models such as the ordered logit or ordered probit model. The case for such models is not very persuasive in the present context, where there are eleven numbered outcomes (from zero to ten) and cardinal interpretations are desired. See Ferrer-i-Carbonell and Frijters (2004) for a comparison of OLS versus ordered response models in this context.

[4]In German: "Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?"

Table 4.1 shows the distribution of the life satisfaction score and summary statistics of the employed variables. The average happiness score lies slightly over 7, where about 50% of the answers are concentrated on the categories 7 and 8. In contrast, only 8% have a life satisfaction score below 5 (the midpoint of the scale). By construction, people in the sample are born between 1904 and 1989. Nearly half of them are women.

4.3.2 Mean

Based on equation (4.2) and the two technical restrictions (4.3) and (4.4), the development of average well-being over the life cycle is analyzed. Figure 4.2 presents the regression results. As discussed in Section 4.2.2, it is not possible to estimate the linear effect of age, cohort and time separately, whereas the linear effect of duration in the panel poses no problem. As argued above, this underidentification is not as severe as other papers suggest, but it is not clear how to report the results. Because this paper focuses on the age effect, the time trend coefficient is restricted to zero (equation (4.4)) and is thus captured in the age and cohort effect curves. The terms "age effect with trend", "cohort effect with trend", respectively "time shock effect" are used to refer to the mapped impacts. Additionally, the estimation results for the isolated linear effects (equation (4.2')) are stated in Table 4.2. The estimated trends are, in contrast to Ree and Alessie (2011), very small. The linear effects of age and time ($\beta^a + \beta^t$), and cohort plus time ($\beta^c + \beta^t$), respectively, are both 0.003 per additional year. The reason for the differences between the results of Ree and Alessie (2011) and this study is the inclusion of duration in the panel as an additional control variable. Because the magnitude of the linear effects is so small, the changes in the graphs would only be minor if the drifts would be excluded.

The age effect with trend can be characterized as U-shaped. However, such a description is somewhat oversimplified. A closer inspection shows that the picture fits nicely to previous research results. There is a small but steady decline in the happiness score between age 20 and 55. After this trough – which coincides with the minimum in a sample of eight European countries found by Blanchflower and Oswald (2009) – average happiness increases strongly until the age of 70. Thereafter, the average score falls sharply. Wunder et al. (2009) and Van Landeghem (2009) have also found a local maximum at age 70. The total magnitude of the effect (0.4) is small and thus in line with, for example, Kassenboehmer and Haisken-DeNew's (2010) doubt about the U-shaped relationship. The effect without linear trend is very similar to the one reported by Ree and Alessie (2011) who also used the GSOEP, albeit for the shorter 1986-2007 period. The reason for the fall at the end of life is likely decreasing health. Explaining the increase after 55 is more difficult and calls for more research.

The cohort profile with trend, displayed in the lower left panel of Figure 2, is slightly increasing. But otherwise no clear or interesting pattern emerges. This is perhaps also due to the low precision of the estimates which renders the interpretation difficult. The time shock profile mirrors the business cycle. There is a distinct peak right after the German reunification. The low in the first decade of the new millennium overlaps with the burst of the ICT bubble. The correlation between the estimated shocks and the GDP growth in the previous year is over 60%.[5] Again, a very similar profile was found by Ree and Alessie (2011). Among all included variables, the time spent in the panel (duration) has clearly the highest impact on reported happiness. The picture suggests a negative linear effect, corroborating the earlier findings of Kassenboehmer and Haisken-DeNew (2010).

4.3.3 Distribution and dispersion

To analyze the distribution of the happiness variable, the baseline model – equation (4.2) together with restrictions (4.3) and (4.4) – is applied separately to each category of the life-satisfaction variable. These linear probability models estimate the effect of age on the probability of having a given life satisfaction score, for example "nine", accounting for year of birth, time shocks, and duration. Figure 4.3 shows the results for the age effect with

[5]Based on calculations using World Bank (2011) data.

trend plus constant, where some categories are combined for simplicity. The results for the controls are not shown. There seem to be no common pattern behind the six curves. The probability of being totally happy (having a value of 10) is steadily decreasing over the life course with a plateau between 60 and 65. This decrease can be made responsible for the downward slope of average life satisfaction between 20 and 55 as well as the fall after 70. The decreasing probability of the highest category implies an offsetting increase for the other categories. The greatest change occurs in the probability of having an 8. The hump shape in the mean curve starting at 55 can largely be ascribed to the temporary decline in the probability of having a 3, 4 or 5, i.e. being rather "dissatisfied". Compared to the rather small absolute changes in average happiness, the changes in the distribution are large. The predicted probability of reporting a 10, for example, decreases by fifteen percentage points over the life cycle.

These findings of decreasing probabilities of low and high life satisfaction scores over the life course imply a steadily shrinking dispersion. To illustrate this, the baseline model is applied to the absolute deviation of the residuals from the mean life satisfaction score regression. Figure 4.4 shows the results. Dispersion is steadily decreasing in age and is smaller for younger cohorts. As discussed, the linear trend of age, cohort and time cannot be disentangled. The most probable explanation for the two decreasing tendencies is a time trend toward more equality, which is mirrored in the graph for the age and the cohort effect. These findings are in line with those of Stevenson and Wolfers (2008) who reported that the dispersion in happiness was shrinking between 1972 and 2006 in the USA, and that happiness is less equally distributed within older cohorts. Compared to age and cohort, duration in the panel and time shocks have only a small effect on dispersion.

4.3.4 Relation in different subgroups

To study heterogeneity in the relationship between age and life satisfaction, the baseline model – equation (4.2) under restrictions (4.3) and (4.4) – is applied to different subgroups.

I consider groupings based on gender and years of education. Among the two, gender is clearly exogenous, whereas education may be affected by self-selection.

Figure 4.5 shows the regression results for men and women. The resemblance of all four curves is quite striking. For both genders, the age curve has a minimum at around 55, followed by a distinct hump shape. The greatest difference in the age effect between the sexes can be observed until 55. Where the curve is clearly decreasing for men, the profile is flatter and perhaps upward sloping between 20 and 30 for women. Further, the positive trend in the cohort effects is stronger for women. Otherwise, there are no noticeable differences between the curves.

To study heterogeneity depending on education, the population is split along the education dimension into six groups of roughly equal size. Figure 4.6 shows the results of the age effect with trend in the baseline model for all types. Two patterns stand out: First, the hump at the end of life can be found in all six groups. Second, the linear trend changes systematically. Life satisfaction for the least educated people is clearly downward sloping. However, the negative trend gets less pronounced as education increases, and the drift for people in the most educated group is even positive. Estimating the baseline model for the mean and including an age-education interaction term confirms the finding that the trend for better educated people is more positive (results not shown in the paper). However, one has to be cautious in interpreting the results. Because people can choose education at the beginning of life, one cannot infer that education causes these different life cycle paths. It is also possible that personality traits, like self-discipline, lead to different life course paths as well as variation in education outcomes.

4.3.5 Finite mixture model

Finite mixture models allow to model heterogeneity depending on unobservable class membership, which does not necessarily depend on observable characteristics like gender or education (a standard reference for finite mixture models is McLachlan and Peel (2000);

for previous happiness applications see Clark et al. (2005) or Bruhin and Winkelmann (2009)). Estimating such a model requires specifying the conditional distribution and not just the mean. Because the specification of the distribution is somewhat arbitrary, two different models are estimated, a linear model with normally distributed error terms and an ordered logit model. Where the first model is a direct extension of the linear model employed earlier, the second has the advantage of respecting the support of the dependent variable $(0, 1, \ldots, 10)$. If a simplified version of the model is estimated, both procedures lead to the same qualitative conclusions. Thus I present only the results of the normal linear model. The log-likelihood contribution of person i, who is represented with T_i observations, in the simplified model is

$$\log\left[\sum_{g=1}^{G}\left(\pi^g \prod_{t=1}^{T_i} \frac{1}{\sigma^g}\phi\left(\frac{y_t - \beta_0^g - \sum_{k=21}^{80}\beta_k^{a,g} I_k^{a_t} - \sum_{k=2}^{26}\beta_k^{d,g} I_k^{d_t}}{\sigma^g}\right)\right)\right], \tag{4.5}$$

where $\phi(.)$ denotes the density function of a standard normally distributed random variable, σ the standard deviation and π^g the probability of belonging to class g. There exist G groups and the subscript g of the parameters indicates that they depend on group membership. Otherwise, the same notation as in equation (4.2) applies.[6] The log-likelihood function is maximized with the EM-algorithm.[7]

Figure 4.7 shows the estimated effect of age on life satisfaction for one to four latent classes. The upper left graph shows the results for the model with only one class. It is evident that the interpretation of a U-shape followed by a hump shape does not change if the time shocks and the cohort variable are excluded. The striking result of the remaining graphs is that the hump shape after 55 can be found in all latent groups, regardless of

[6]The estimated model imposes the restrictions that the time and cohort variables have no effect. This facilitates convergence and is compatible with the theoretical independence between the time shocks and age, and the empirical finding that the cohort effect has no systematic effect on mean life satisfaction (section 4.3.2). Further, people age 20 polled the first time (hence duration=1) are defined as the base category.

[7]The R-program FlexMix by Bettina Gruen and Friedrich Leisch (2008) was used to estimate the linear finite mixture model.

the number of allowed classes. The trends over the life cycle, however, differ between the groups. The largest group is always the one with no clear drift.

4.3.6 Individual patterns

The study of the relation between age and life satisfaction on the individual level is complicated by three factors. First, the maximum length one can follow a person is 26 years and it is therefore not possible to study satisfaction over the whole life cycle for any one individual. Second, the variance of the error term is large relative to the expected changes of the mean happiness score. At the individual level, the smallest possible change of the dependent variable is one point. This exceeds the maximal average effect found over the life cycle. Third, duration in the panel has a large effect. But on the individual level, it is not possible to disentangle the duration from the age effect in a credible manner.

The empirical inquiry at the individual level starts therefore by studying the fraction of people at a certain age, who report at that age a larger (or smaller) happiness than at any other time over the previous ten years. This restricts the analysis to individuals who are observed for at least eleven years. The share of people in the population experiencing a minimum (maximum) is probably overestimated (underestimated), because this analysis ignores the duration effect. However, it is still possible to determine whether the finding of a U-shape followed by a hump shape prevails at the individual level. Figure 4.8 shows the results. First a short explanation how to read the graph: A value of 20% at the age of thirty means that one out of five people, who are at least observed between age twenty to thirty, experiences a minimum at thirty in this period (the minimum has not to be unique). The confidence intervals are not shown for ease of readability (but each estimate is based on more than 1000 observations). The fraction of minimums almost always exceeds the fraction of maximums. The obvious explanation is the neglected duration effect. There is no clear trough at 55 but the fraction of minimums decreases and the proportion of maximums goes up after this age. These small changes are more than offset by the trends after age 68. The fraction of minimums is nearly exploding, and the fraction of maximums shrinks.

Because the U-shape hypothesis and the corresponding trough have received much attention in the literature, the distribution of minimums for people who are observed between age 48 and 62 are shown in Figure 4.9, these are about 1000 individuals. The distribution is nearly uniform, with a slight increase with age. This slight trend can again be explained by the neglected negative duration effect. The general pattern suggests that only a small fraction of the individuals reach the minimum at exactly 55.

To further study heterogeneity at the individual level, a separate model is estimated for each individual and each possible interval of length eleven (thus, in general, more than one model per individual). To keep it simple, the model consists of two linear age terms, one for the first and one for the second half of the interval. Each regression is then characterized as hump shaped (if the first linear term is positive and the second one is negative), U-shaped (if the first linear term is negative and the second one is positive), increasing or decreasing (depending on if both terms are positive or negative). If well-being does not systematically change over the life course, the four curves should be flat. But this is obviously not the case, as can be seen in Figure 4.10. The picture largely confirms the finding that the hump shape after 55 is the dominant pattern. The fractions are nearly stable until 50 where the U-shape and the increasing patterns start to gain shares. At 55, the fraction of U-shape types reaches a maximum and the fraction of hump shape types a minimum. Shortly after 60, the fraction of decreasing types becomes more and more important.

4.4 Conclusion

This paper studies the relationship between age and self-reported well-being not just at the average level, as customary in prior research, but also at the individual level, analysing the differences between individual life cycle paths. The inquiry of heterogeneity at the individual level, while of substantial interest, is hampered by the rough measure of life satisfaction and the related high volatility in individual life cycle paths, as well as by the strong duration effect. Nevertheless, it is safe to conclude that a life cycle pattern with two turning points is prominent at the individual level as well.

Insights into the average evolution of life satisfaction and group differences between individual life cycle paths are more robust. Mean life satisfaction is steadily declining between 20 and 55. After this low, happiness increases strongly until the age of 70, where it starts to fall sharply. The driving force behind the hump shape after 55 is the temporarily diminishing share of rather unsatisfied people. The mild downward trend, on the other side, is mostly due to the falling probability of the "completely satisfied" category. The dispersion in well-being is decreasing in age and year of birth. An explanation of this pattern is the time trend toward more equality.

While the happiness trend over the life course differs between groups of people, whereas the form of the relationship (thus the deviation from the linear trend), is rather stable, namely a hump shape after 55 with a peak at around 70. This conclusion is based on the regression for different education groups as well as the results of the finite mixture model. Gender differences are minor.

Further research should concentrate on the channels through which age affects life satisfaction. Because of the large reporting effect of duration in the panel, repeated cross sectional data are probably most suitable for such a task. The gain from panel data, namely the potential study of individual life cycle paths, can hardly offset this drawback.

References

Blanchflower, D. G., and Oswald, A. J. (2004): "Well-Being over Time in Britain and the US", *Journal of Public Economics*, **88**, 1359-1386.

Blanchflower, D. G., and Oswald, A. J. (2008): "Is Wellbeing U-shaped over the Life Cycle", *Social Science and Medicine*, **66**, 1733-1749.

Blanchflower, D. G., and Oswald, A. J. (2009): "The U-Shape without Controls: A response to Glenn ", *Social Science and Medicine*, **69**, 486-488.

Bruhin, A., and Winkelmann, R. (2009): "Happiness Functions with Preference Interdependence and Heterogeneity: The Case of Altruism within the Famil", *Journal of Population Economics*, **22**, 1063-1080.

Clark, A. E. (2007): "Born To Be Mild? Cohort Effects Don't (Fully) Explain Why Well-Being Is U-Shaped in Age", IZA Discussion Paper No. 3170.

Clark, A. E., Diener, E., Geogellis, Y., and Lucas, R. E. (2008): "Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis", *The Economic Journal*, **118**, 222-243.

Clark, A. E., Etilé, F., Postel-Vinay, F., Senik, C., and Van der Straeten, K. (2005): "Heterogeneity in reported well-being: Evidence from twelve European countries", *The Economic Journal*, **115**. 118 – 132.

Easterlin, R. A. (2006): "Life cycle happiness and its sources: Intersections of psychology, economics and demography", *Journal of Economic Psychology*, **27**, 463-482.

Easterlin, R. A., and Sawangfa, O. (2007): "Happiness and Domain Satisfaction: Theory and Evidence", IZA Discussion Paper No. 2584.

Ferrer-i-Carbonell, A., and Frijters, P. (2004): "How important is methodology for the estimates of the determinants of happiness?", *Economic Journal*, **114**, 641-659.

Frijters, P., and T. Beatton (2011): "The mystery of the U-shaped relationship between happiness and age", NCER Working Paper Series 26R.

Fujita, F., and Diener, E. (2005): "Life Satisfaction Set Point: Stability and Change", *Journal of Personality and Social Psychology*, **88**, 158-164.

Fischer, J. AV. (2009): "Happiness and age cycles – return to start…", MPRA paper No. 15249.

Glenn, N. (2009): "Is the apparent U-shape of well-being over the life course a result of inappropriate use of control variables? A commentary on Blanchflower and Oswald (66: 8, 2008, 1733-1749)", *Social Science and Medicine*, **69**, 481-485.

Gruen, B., and Leisch, F. (2008): "FlexMix Version 2: Finite mixtures with concomitant variables and varying and constant parameters", *Journal of Statistical Software*, **28**, 1-35.

Gwozdz, W., and Sousa-Poza, A. (2010): "Ageing, Health and Life Satisfaction on the Oldest Old: An Analysis for Germany", *Social Indicator Research*, **97**, 397-417.

Kassenboehmer, S., and Haisken-DeNew J. P. (2010): "Heresy or Enlightenment? The Wellbeing Age U-Shape Effect is Really Flat!" Unpublished manuscript.

McKenzie, D. J. (2006): "Disentangling Age, Cohort and Time Effects in the Additive Model", *Oxford Bulletin of Economics and Statistics*, **68**, 473-495.

McLachlan, G., and Peel, D. (2000): *Finite Mixture Models*. Wiley, New York.

Mroczek, D. K., and Spiro A. (2005): "Change in Life Satisfaction during Adulthood: Findings From the Veterans Affaires Normative Aging Study", *Journal of Personality and Social Psychology*, **88**, 189-202.

Ree, J., and Alessie, R. (2011): "Life Satisfaction and Age: Dealing with Underidentification in Age-Period-Cohort Models", *Social Science and Medicine*, **73**, 177-182.

Schilling, O. (2006): "Development of life satisfaction in old age: another view on the "paradox"", *Social Indicator Research*, **75**, 241-271.

Stone, A. A., Schwartz, J. E., Broderick, J. E. and Deaton, A. (2010): “A snapshot of the age distribution of psychological well-being in the United States”, *Proceedings of the National Academy of Sciences*, **107**, 9985-9990.

Stevenson, B., and Wolfers, J. (2008): “Happiness Inequality in the United States”, *Journal of Legal Studies*, **37**, 33-78.

Van Landeghem, B. (2009): “The Course of Subjective Well-Being over the Life Cycle”, *Journal of Applied Social Science Studies*, **129**, 261-267.

Van Landeghem, B. (2011): “A Test for the Convexity of Human Well-Being over the Life Cycle: Longitudinal Evidence from a 20–Year Panel”, *Journal of Economic Behaviour and Organization*, forthcoming.

Winkelmann, L., and Winkelmann, R. (1998): “Why are the unemployed so unhappy? Evidence from panel data”, *Economica*, **65**, 1-15.

Wunder, C., Wiencierz, A., Schwarze, J., Küchenhoff, H., Kleyer, S., and Bleninger, P. (2009): “Well-being over the life span: evidence from British and German longitudinal data”, IZA Discussion Papers No. 4155.

World Bank (2011): *World Development Indicators*.

Figure 4.1: Illustration of the aggregation problem

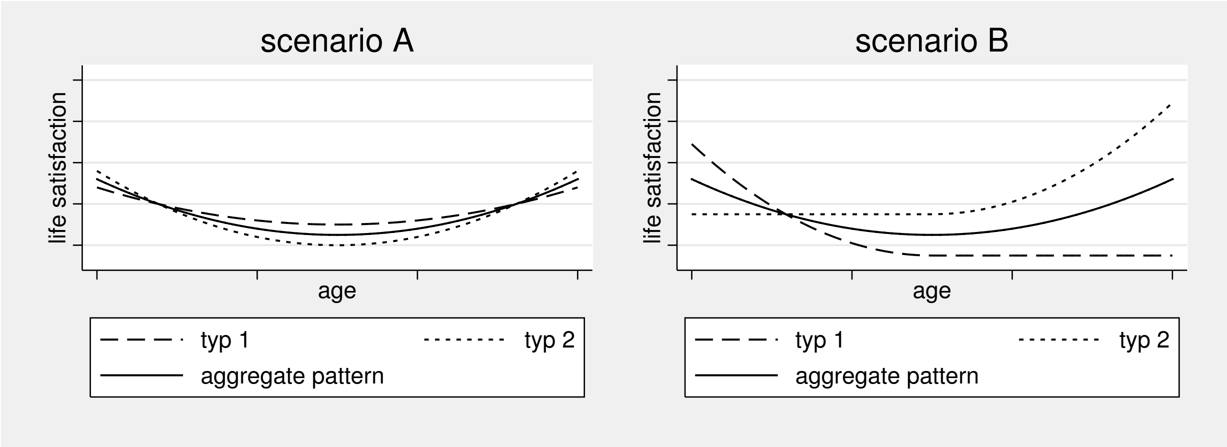
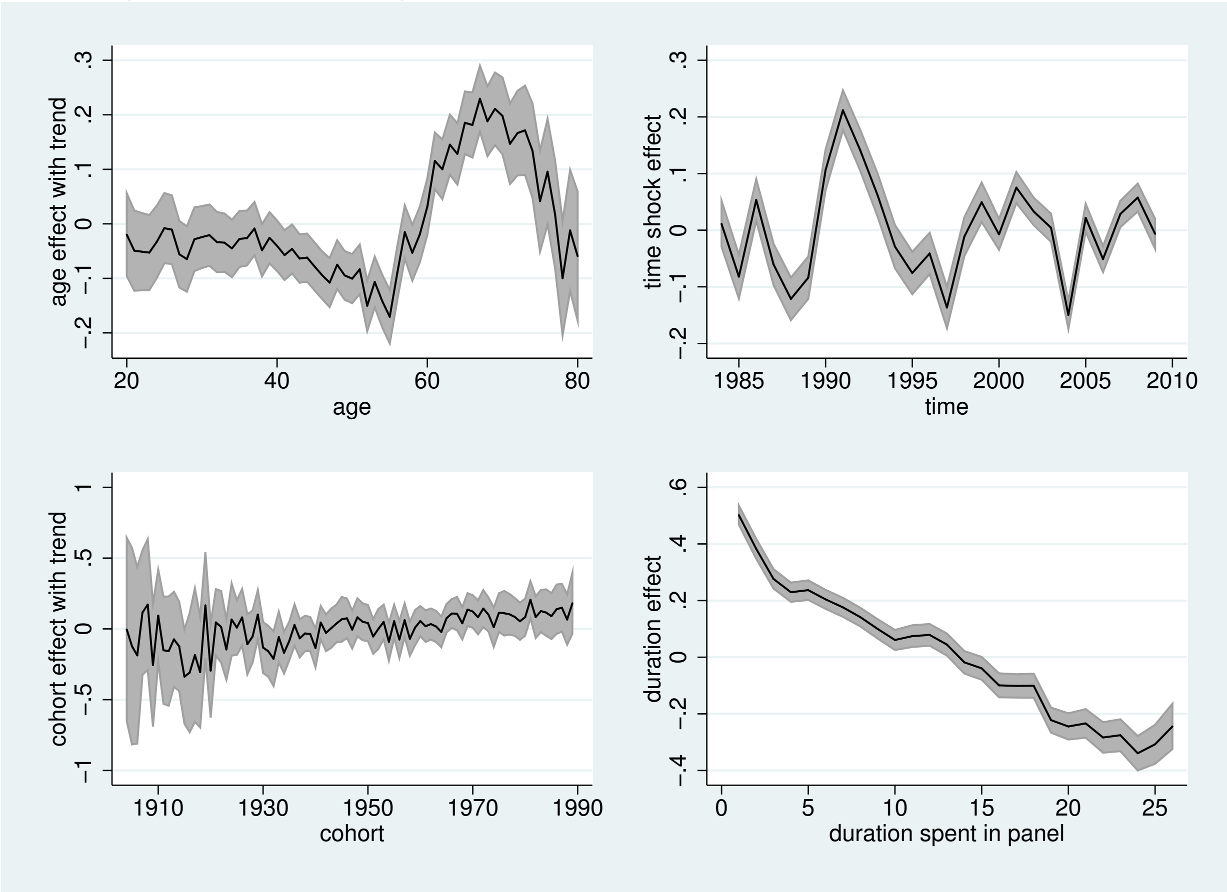
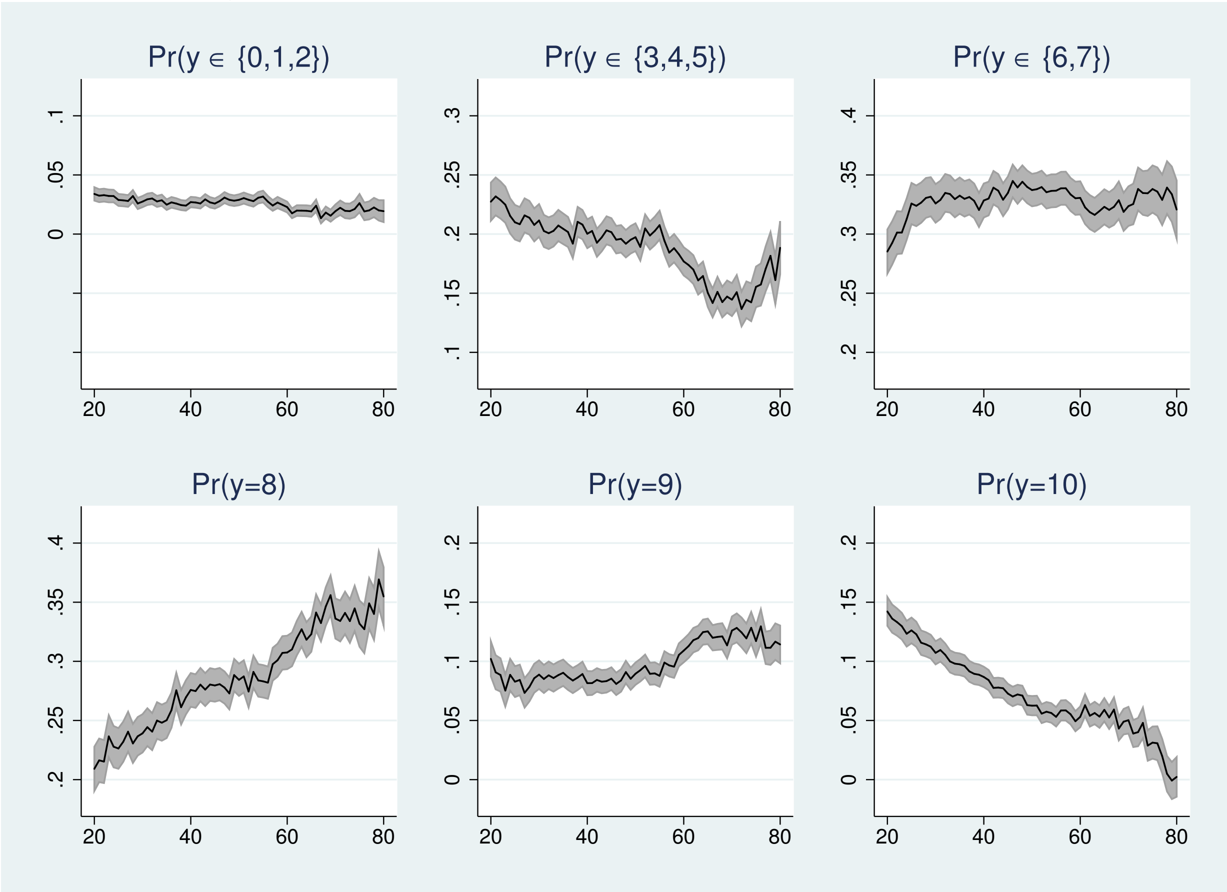


Figure 4.2: Effect of age, cohort, time and duration on mean life satisfaction



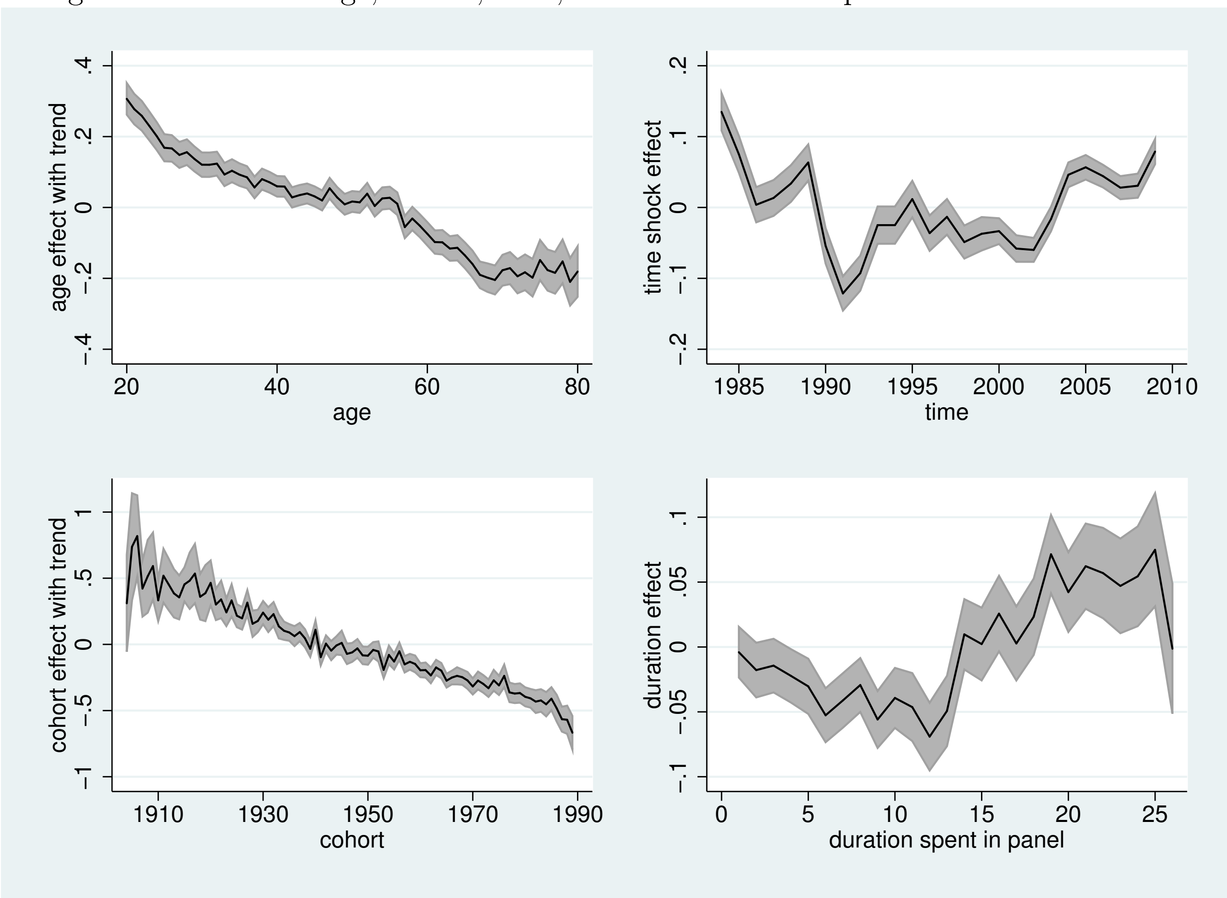
Notes: The black line depicts the estimates of equation (4.2) under the restrictions (4.3) and (4.4). The gray area indicates the 95% confidence intervals computed based on cluster robust standard errors.

Figure 4.3: Distribution of life satisfaction over the life course



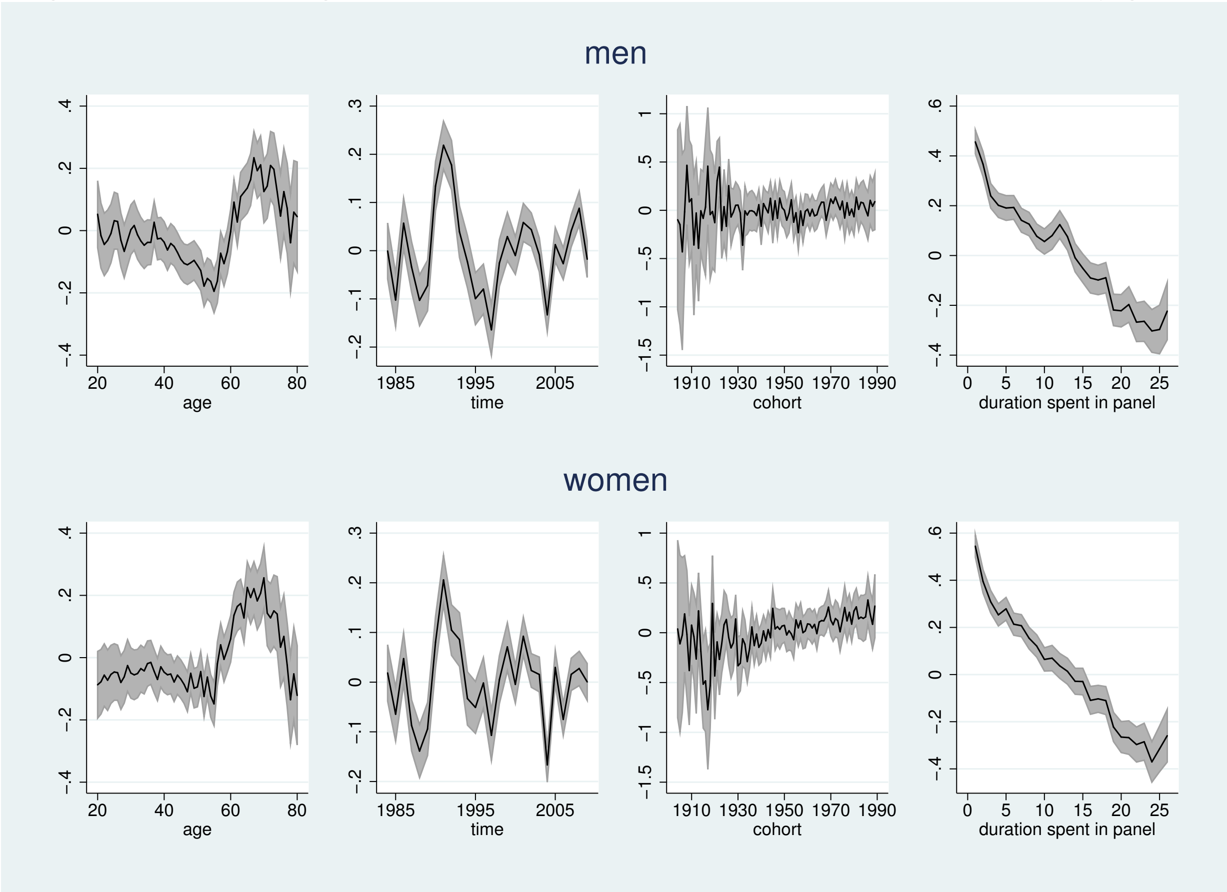
Notes: Results based on linear probability models. The black line depicts the estimated age effect with trend plus constant (equation (4.2) under restrictions (4.3) and (4.4)). The gray area indicates the 95% confidence interval computed based on cluster robust standard errors.

Figure 4.4: Effect of age, cohort, time, and duration on dispersion of life satisfaction



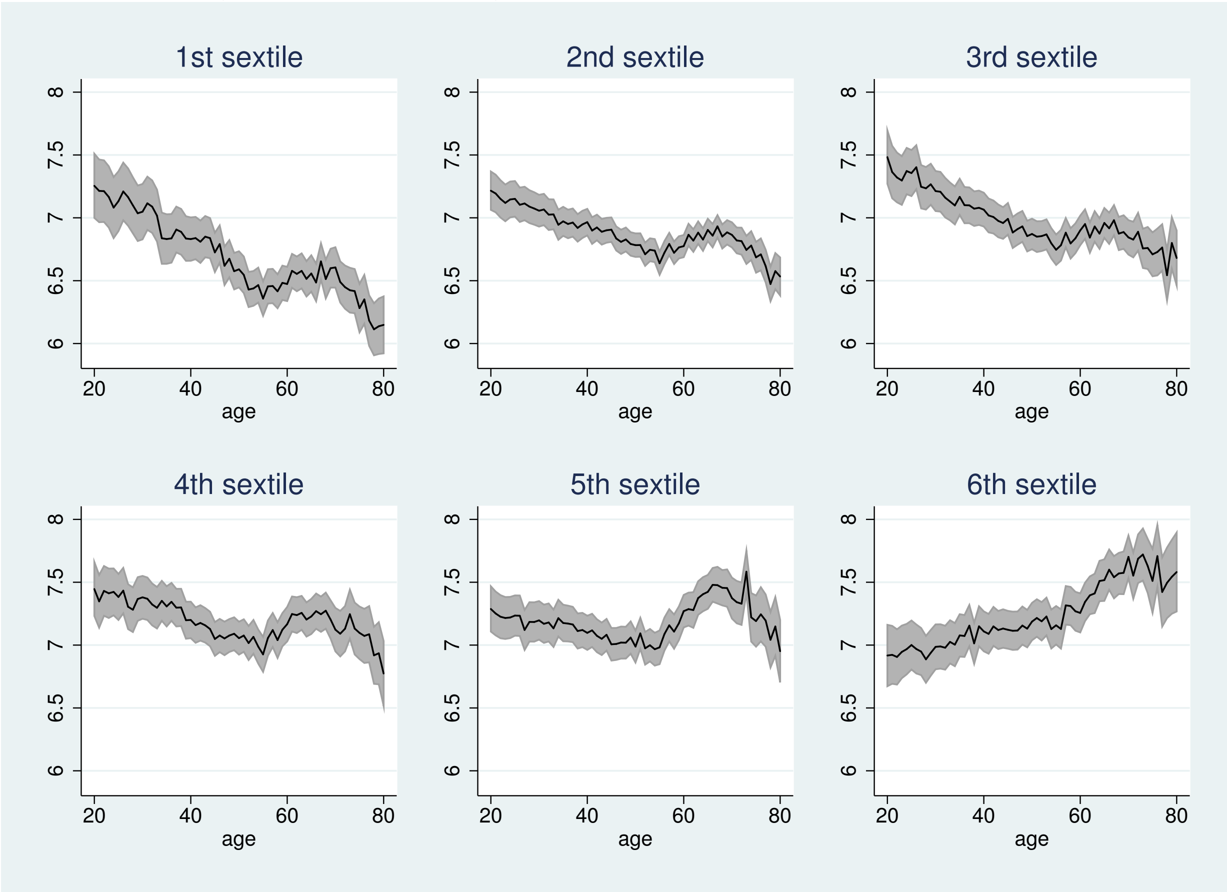
Notes: The dependent variable is the absolute deviation of the residuals from the mean life satisfaction regression. The regression is based on equation (4.2) under restrictions (4.3) and (4.4). The gray area indicates the 95% confidence intervals computed based on cluster robust standard errors.

Figure 4.5: Effect of age, cohort, time, and duration on mean life satisfaction by gender



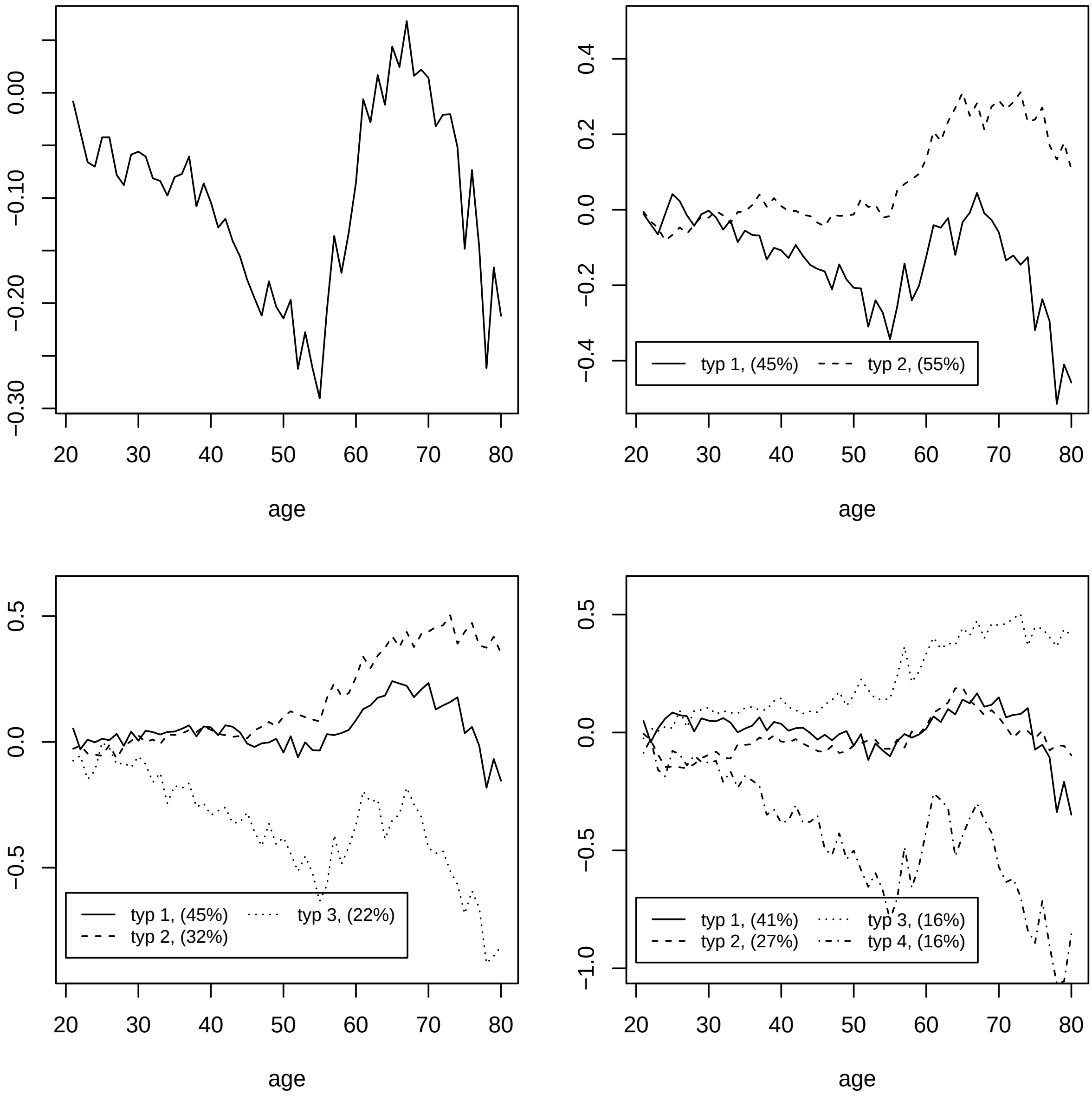
Notes: The upper graph shows the results for men, the lower graph shows the results for women. Otherwise, legend of Figure 4.2 applies.

Figure 4.6: Effect of age on mean life satisfaction by education



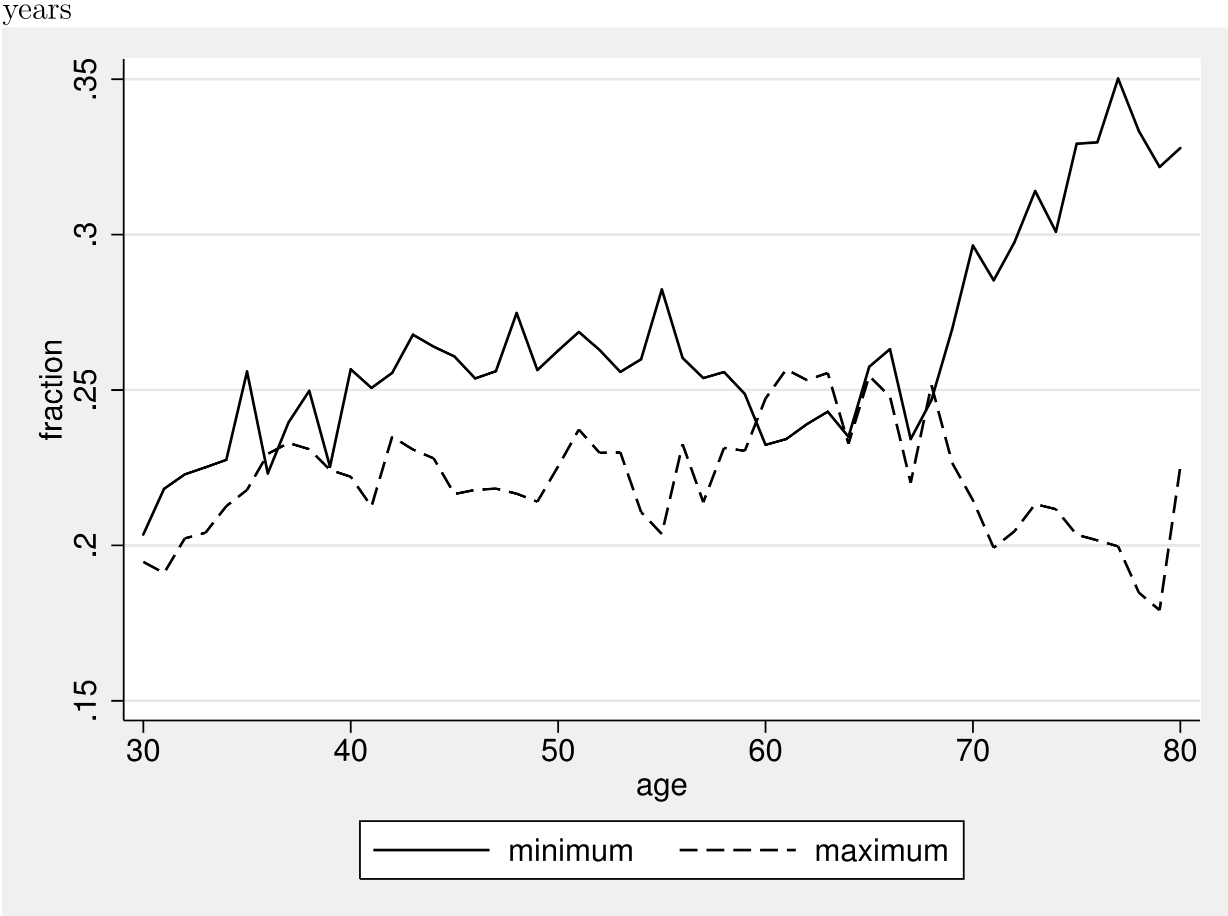
Note: The black line depicts the estimated age effect with trend (equation (4.2) under restrictions (4.3) and (4.4)) plus constant for different education sextiles. The gray area indicates the 95% confidence interval computed based on cluster robust standard errors.

Figure 4.7: Finite mixture model: Effect of age on mean life satisfaction



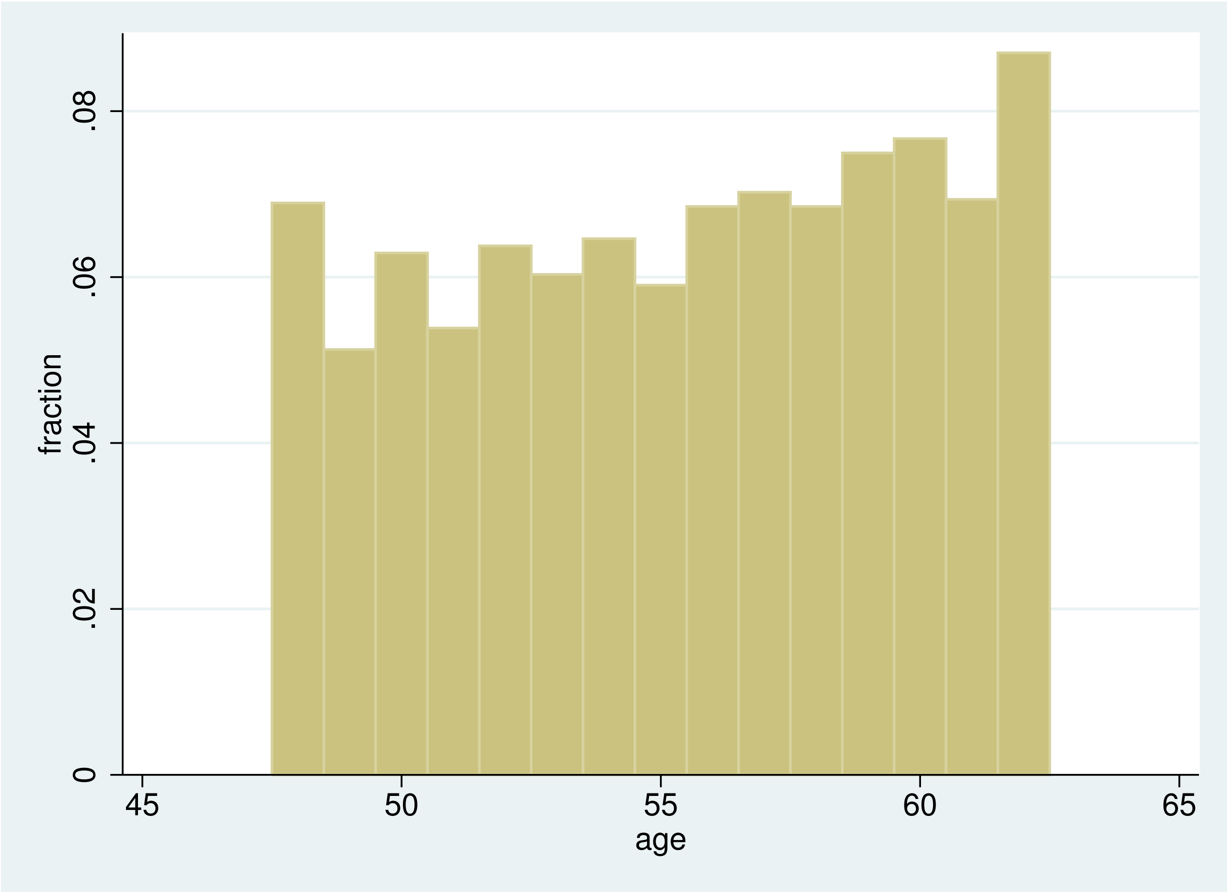
Notes: Estimated age effect on life satisfaction and estimated fraction of each class in linear finite mixture models (with normally distributed error terms) with up to four latent classes (equation (4.5)).

Figure 4.8: Fraction of people reaching a minimum or maximum compared to the last ten years



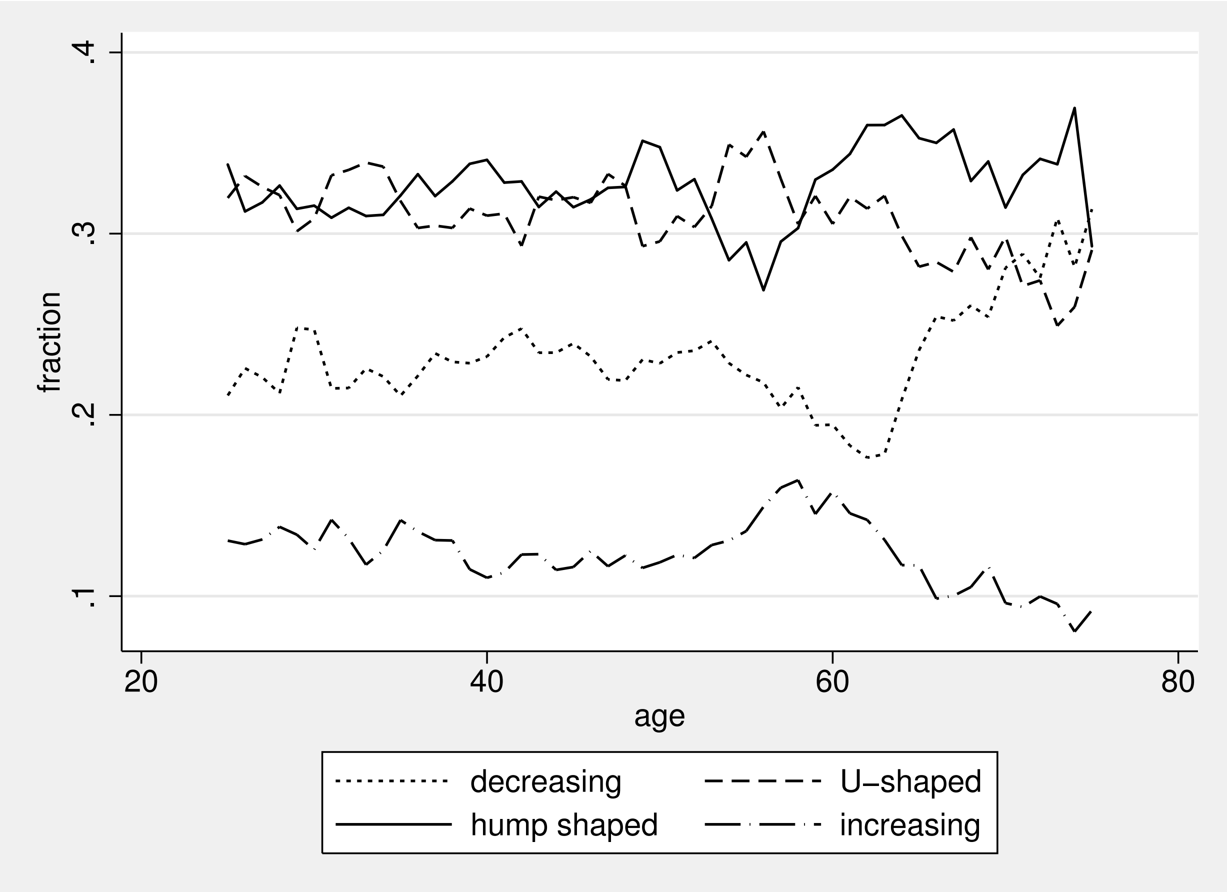
Note: The continuous (dotted) line indicates the fraction of people, who reach at this specific age the lowest (highest) life satisfaction level compared to the last ten years where the minimum (maximum) has not to be unique.

Figure 4.9: Distribution of observed minimums



Note: Distribution of minimums for people observing between age 48 and 62. Number of people: 979, number of minimums: 2321 (minimum has not to be unique).

Figure 4.10: Distribution of evolution patterns



Note: Fraction of evolution patterns between 25 and 75. In a first step, for every person and all possible intervals of length eleven, a simple model with two linear terms, one for the first and one for the second part, is estimated: $y = \beta_0 + \beta_1(a\mathrm{I}(t-c-a < 0)) + \beta_2(a\mathrm{I}(t-c-a > 0)) + \epsilon$ if $|t-c-a| \leq 5$, where the standard notation of the paper applies. In a second step, the results are classified ($\beta_1 < 0$ $\beta_2 < 0 \rightarrow$ decreasing, $\beta_1 < 0$ $\beta_2 > 0 \rightarrow$ U-shaped, $\beta_1 > 0$ $\beta_2 < 0 \rightarrow$ hump shaped, $\beta_1 > 0$ $\beta_2 > 0 \rightarrow$ increasing) and the fraction of each type is computed. Reading example: Between 30 and 40 (thus value at age 35), about twenty percent of the individual patterns can be described as decreasing.

Table 4.1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.
life satisfaction	7.138	1.808	0	10
life satisfaction = 0	0.005		0	1
life satisfaction = 1	0.004		0	1
life satisfaction = 2	0.011		0	1
life satisfaction = 3	0.023		0	1
life satisfaction = 4	0.032		0	1
life satisfaction = 5	0.110		0	1
life satisfaction = 6	0.103		0	1
life satisfaction = 7	0.211		0	1
life satisfaction = 8	0.309		0	1
life satisfaction = 9	0.123		0	1
life satisfaction = 10	0.069		0	1
year	1998.0	7.6	1984	2009
age	45.6	15.7	20	80
year of education	11.5	2.6	7	18
female	0.514		0	1
cohort (year of birth)	1952.4	16.5	1904	1989
duration (year in panel)	8.0	6.2	1	26

Notes: Data from GSOEP. The used sample consists of 38,197 different individuals, 304,856 person-year observations, living in (former) West-Germany.

Table 4.2: Estimated trends in average life satisfaction

	coeff	std. err.
age + time	0.0029	(0.0011)
cohort + time	0.0029	(0.0011)
duration	-0.0295	(0.0015)
constant	6.9753	(0.0180)
Observations	304,856	
Individuals	38,197	

Notes: The table shows the estimation results for (4.2'). Standard errors in parentheses are corrected for clustering at the individual level. Not shown are nonlinear components of the age profile (60 dummies), of the cohort profile (86 dummies), of the year profile (26 dummies), and of the duration profile (26 dummies). These are together with the trends displayed in Figure 4.2.

Chapter 5

Occurrence dependence and zero-inflation in count data models

This chapter is joint work with Rainer Winkelmann.

Acknowledgements: Valuable comments by Karim Chalak, Stefan Hoderlein, Maximilian Kasy, as well as seminar participants at Harvard University and Boston College are gratefully acknowledged.

5.1 Introduction

Count data models are used in situations where events occur repeatedly over time whilst the timing of events is not recorded and thus unbeknown to the investigator. Only the total number of events in a fixed time interval is observed. Examples are the number of doctor visits during a quarter, the annual number of patent applications, or the number of times, an investor goes online to look up his portfolio valuation during a week.

The statistical properties of count data are inherited from the properties of the underlying stochastic process that determines the timing of events. Heckman and Borjas (1980) characterize continuous time multiple spell processes in terms of three dimensions: occurrence dependence, duration dependence, and unobserved heterogeneity. Occurrence dependence means that the mere occurrence of an event alters the probability of future events. A special case of occurrence dependence arises if there is a single switch in the rate after the first event. Since the timing of that first event is random, such a process can be said to possess a "stochastic hurdle".

If events occur randomly over time, without occurrence dependence, duration dependence, or unobserved heterogeneity, then the number of events during a fixed time interval is Poisson distributed. Departures along any one of the three dimensions lead to a different count data model ("non-Poissonness"). In the past, a considerable amount of research has been devoted to the consequences of unobserved heterogeneity (Hausman, Hall and Griliches, 1984, Cameron and Trivedi, 1986) and duration dependence (Winkelmann, 1995, McShane et al., 2008). Occurrence dependence in general, and a stochastic hurdle process in particular, have been left unexplored by the earlier econometrics literature, although there are some precursors in biometrics (see, e.g., Faddy, 1997, and Janardan, 1980).

This neglect is unfortunate, since a stochastic hurdle model is likely to be very useful in practice as it can provide a natural explanation for extra zeros in count data, a phenomenon that is very frequently encountered in econometric count data applications (see, for example Pizer and Prentice, 2011; Sari, 2009; Sarma and Simpson, 2006; Yen, Tang and Su, 2001;

Chang and Trivedi, 2003; Street, Jones and Furuta, 1999). The two existing models for addressing the presence of extra-zeros, the zero-inflated count data model (see e.g. Lambert, 1992) and the fixed-hurdle count data model (see e.g. Mullahy, 1986) lack credible data generating processes. While the zero-inflated count model makes the extreme assumption that the population can be split into two sub-populations one of whom never experiences the event, regardless of the length of time, the fixed-hurdle model cannot be reconciled with, or interpreted in terms of, a meaningful underlying stochastic count process. The usefulness of these existing models therefore remains limited. While they may give approximately valid answers for questions regarding mean effects, they cannot be used to establish certain policy counterfactuals, such as the effect on the distribution of counts of extending the time period from T to $2T$, or the extensive margin effect for observation units who cross the stochastic hurdle because of the policy change, i.e., those who have a potential outcome of zero in absence of the policy change and a positive outcome otherwise. These questions require specifying the underlying structural count process, and deriving the count distribution from there.

The new model in this paper is based on a count process with occurrence dependence, where the occurrence rate in the underlying event-generating process until the first event can differ from that applying to subsequent events. The resulting *stochastic hurdle model* generates zero-inflation if the rate is initially low and increases after the first occurrence. Importantly, in such a framework, the probability of zero and the distribution of positive outcomes cannot be treated independently: Variation in the first rate systematically affects the expected arrival time of the first event, and hence the duration for which the process is in the second state. This *time effect* is accounted for in neither the fixed-hurdle model nor in zero-inflated model. Moreover, the new model allows to address the effect of exposure time in a theory-consistent way, and it overcomes shortcomings of previous decompositions into extensive and intensive margin effects.

The paper proceeds as follows. In the next section, we present the standard models for zero-inflated count data. In section 5.3, we derive the stochastic hurdle model and discuss its properties. The new approach is used, in section 5.4, to estimate the effect of a health care reform that took place in Germany in 1997 on the number of quarterly visits to a physician, based on survey data from the German Socio-Economic Panel. Section 5.5 concludes.

5.2 Modeling zero-inflation

The standard Poisson model has probability function

$$\Pr(Y = k) = \exp(-\lambda)\lambda^k/k!, \quad k = 0, 1, 2, \ldots$$

If λ follows a gamma distribution, the resulting marginal probability function for Y is negative binomial (e.g. Winkelmann, 2008). The negative binomial distribution has, for a given mean, a larger variance than the Poisson distribution (overdispersion). It also has a higher probability of a zero.

In many applications, the extra zeros (relative to the Poisson model) generated by the negative binomial model are insufficient to account for the full amount of zeros in the data. Moreover, one often has a substantive (structural) interest to treat the zero-generating process separately from the process for strictly positive outcomes, which requires different sets of parameters. There are two standard ways of doing this. In the fixed-hurdle (FH) model (Mullahy, 1986)

$$\Pr(Y = k) = \begin{cases} \phi & \text{for } k = 0 \\ (1-\phi)\dfrac{f_P(k)}{1 - f_P(0)} & \text{for } k = 1,\ 2,\ 3, \ldots \end{cases} \tag{5.1}$$

where $f_P(k)$ denotes the Poisson probability function. In the zero-inflated (ZIP) model (Lambert, 1992)

$$\Pr(Y = k) = \begin{cases} \phi + (1-\phi) f_P(0) & \text{for } k = 0 \\ (1-\phi) f_P(k) & \text{for } k = 1,\ 2,\ 3, \ldots \end{cases} \tag{5.2}$$

A key difference between the two approaches is the origin of zeros: while there is a single type of zero under the FH assumption, there are two types in the ZIP model, sometimes referred to as "strategic" versus "incidental". Which of the two assumptions is more plausible depends on the specific application. For example, Pohlmeier and Ulrich (1995) have argued that a hurdle model can well represent the demand for doctor visits, where a first decision to contact a GP might be followed by a number of re-appointments or referals to specialists that are subject to a different mechanism. Unfortunately, the hurdle model ignores the timing dimension, i.e., the difference it makes whether the first contact was made earlier or later during the observation period. Our approach, by contrast, directly addresses the dynamics of the count process.

5.3 Occurrence dependence and stochastic hurdle

There is a fundamental relationship between a multiple-spell stochastic process and the resulting number of counts (see e.g. Winkelmann, 1995). Denote the arrival time of the k-th event by ϑ_k. Let $N(T)$ represent the total number of events between 0 and T. The probability that at most $k-1$ events occurred before T equals the probability that the arrival time of the k-th event is greater than T:

$$\Pr(N(T) < k) = \Pr(\vartheta_k > T) \tag{5.3}$$

Moreover

$$\begin{aligned} \Pr(N(T) = k) &= \Pr(N(T) < k+1) - \Pr(N(T) < k) \\ &= \Pr(\vartheta_{k+1} > T) - \Pr(\vartheta_k > T) \\ &= F_k(T) - F_{k+1}(T) \end{aligned} \tag{5.4}$$

where F_k is the cumulative density function of ϑ_k and it is understood that $F_0(T) = 1$. Thus, the count distribution is fully determined once the distribution functions of arrival times are specified for all $k \geq 1$.

Let $\tau_k = \vartheta_k - \vartheta_{k-1}$ denote the *interarrival* times. Suppose τ_1 (the arrival time of the first event) is exponentially distributed with rate λ_1. All subsequent interarrival times are also exponential, but with rate λ_2, where $\lambda_1 \neq \lambda_2$. There is a discrete and one-time change in the underlying rate. Except for this change (i.e. occurrence dependence), spells are assumed to be independent. Thus, the arrival time of the first event has density function

$$f(\vartheta_1) = \lambda_1 \exp(-\lambda_1 \vartheta_1)$$

The density function of the arrival time of the second event is obtained as

$$f(\vartheta_2) = \int_0^{\vartheta_2} \lambda_1 \exp(-\lambda_1(\vartheta_2 - t))\lambda_2 \exp(-\lambda_2 t)dt = \frac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2}[\exp(-\lambda_1\vartheta_2) - \exp(-\lambda_2\vartheta_2)]$$

(see Feller, 1977). The arrival time density of the k-th event can be obtained in a similar way. Assuming constant and independent renewals from the first event onwards, the distribution of the interarrival time between the 1st and the k-th event is known to be of Erlang form (see e.g. Winkelmann, 1995), and therefore

$$f(\vartheta_k) = \int_0^{\vartheta_k} \lambda_1 \exp(-\lambda_1(\vartheta_k - t))\frac{\lambda_2^{k-1}}{\Gamma(k-1)} t^{k-2} \exp(-\lambda_2 t)dt$$

Solving the integral and applying (5.4) yields a corresponding (modified) count data model with zero-inflation (as long as $\lambda_2 > \lambda_1$).

An alternative derivation works directly with probabilities. Suppose as before that the first event occurs at time $\vartheta_1 = t$, and that $k-1$ events occur between t and T. For independent exponentially distributed interarrival times with two different rates, λ_1 and λ_2, the joint probability is $\Pr(Y = k, \vartheta_1 = t) = \lambda_1 \exp(-\lambda_1 t) f_P(k-1; \lambda_2(T-t))$. Integrating over the unobserved t, we obtain the probability function

$$\Pr(Y = k; \lambda_1, \lambda_2, T) = \int_0^T \lambda_1 \exp(-\lambda_1 t) f_P(k-1; \lambda_2(T-t))dt$$

The integral on the right is equal to

$$f_{SH}(k; \lambda_1, \lambda_2, T)$$
$$= \begin{cases} \exp(-\lambda_1 T) & \text{for } k = 0 \\ \dfrac{\lambda_1 T(\lambda_2 T)^{k-1} \exp(-\lambda_1 T)}{(\lambda_2 T - \lambda_1 T)^k} \left[1 - \displaystyle\sum_{j=0}^{k-1} \dfrac{\exp(-\Delta)\Delta^j}{j!}\right] & \text{for } k = 1, 2, 3, \ldots \end{cases} \tag{5.5}$$

with $\Delta = \lambda_2 T - \lambda_1 T$. We denote this probability function by f_{SH} for "stochastic hurdle". The process randomly switches from one state (state 1 with intensity λ_1) to a second state (state 2 with intensity λ_2). If T is the same for everyone, it can be set equal to 1 without loss of generality. Details of the derivation can be found in Appendix A. If $\lambda_2 > \lambda_1$ the term in squared brackets equals $1 - F_p(k-1, \lambda_2 T - \lambda_1 T)$ where F_p is the cumulative distribution function of the Poisson distribution.

Figure 5.1 shows plots of the probability functions for the Poisson distribution, the SH model, the FH model and the ZIP model. The distributions are standardized such that the mean is 2.3 in all cases. The probability of a zero is $\exp(-2.3) = 10\%$ under the Poisson assumption, whereas it is 27% in the other three models. Thus, there is massive zero-inflation. Figure 1 illustrates the key difference between the standard zero-inflated models on the one hand, and the stochastic hurdle model on the other: whereas the conditional-on-positives distributions of the former two are scaled versions of a Poisson distribution (with mean adjusted so that the overall mean is 2.3), the stochastic hurdle model is more spread out: the probability of a one exceeds that of the FH and ZIP models, and the same holds for outcomes in the right tail of the distribution.

5.3.1 Expected value

Using (5.5) we obtain

$$
\begin{aligned}
\mathrm{E}_{SH}(y;\lambda_1,\lambda_2,T) &= \sum_{y=1}^{\infty} y \int_0^T f_P(y-1;\lambda_2(T-t))f(t;\lambda_1)dt \\
&= \int_0^T \sum_{y=0}^{\infty}(y+1)f_P(y;\lambda_2(T-t))f(t;\lambda_1)dt \\
&= \int_0^T [\lambda_2(T-t)+1]f(t;\lambda_1)dt \\
&= \Pr(y>0;\lambda_1)+\lambda_2\mathrm{E}(T-t;\lambda_1) \qquad (5.6)
\end{aligned}
$$

This is a key result: The expectation is the sum of the probability of passing the stochastic hurdle, plus the state 2 rate times the expected duration in state 2. The equation shows that λ_1 affects the overall mean through two separate channels. First, it affects the probability of crossing the hurdle, and second, it affects the expected duration spent in the second state. This distinction is absent in the FH model, where the expectation is given by

$$
\begin{aligned}
\mathrm{E}_{FH}(y;\lambda_1,\lambda_2,T) &= \Pr(y>0;\lambda_1,T)\mathrm{E}(y|y>0;\lambda_2,T) \\
&= [1-\exp(-\lambda_1 T)]\,\frac{\lambda_2 T}{1-\exp(-\lambda_2 T)} \qquad (5.7)
\end{aligned}
$$

One can show that $\partial\mathrm{E}_{SH}(y)/\partial\lambda_1 > \partial\mathrm{E}_{FH}(y)/\partial\lambda_1$ if the two models have the same expected value and the same fraction of zeros.

For the SH model, the expected time spent in the first state is

$$
\begin{aligned}
\mathrm{E}(t;\lambda_1,T) &= \Pr(y=0)T+\int_0^T \exp(-\lambda_1 t)\lambda t\,dt \\
&= \exp(-\lambda_1 T)T + 1/\lambda_1 - (T+1/\lambda_1)\exp(-\lambda_1 T) \\
&= 1/\lambda_1(1-\exp(-\lambda_1 T))
\end{aligned}
$$

and the expected time spent in the second state is therefore

$$
\mathrm{E}(T-t;\lambda_1,T) = T - 1/\lambda_1(1-\exp(-\lambda_1 T)) \qquad (5.8)
$$

The first derivative of $\mathrm{E}(T-t;\lambda_1)$ with respect to λ_1 is positive, because a higher rate of the state 1 process tends to reduce the time spent in that state, increasing the time left for state 2. Inserting (5.8) into (5.6), we obtain

$$\begin{aligned}
\mathrm{E}_{SH}(y;\lambda_1,\lambda_2,T) &= \Pr(y>0;\lambda_1,T)+\lambda_2\mathrm{E}(T-t;\lambda_1) \\
&= [1-\exp(-\lambda_1 T)]+\lambda_2[T-1/\lambda_1(1-\exp(-\lambda_1 T))] \\
&= \lambda_2 T+(1-\lambda_2/\lambda_1)[1-\exp(-\lambda_1 T)] \qquad (5.9)
\end{aligned}$$

As required, the expected value reduces to the Poisson mean when $\lambda_1 = \lambda_2$. The expected value is greater than λ_2 when $\lambda_1 > \lambda_2$, and smaller otherwise.

5.3.2 Observed heterogeneity

In cross-sectional or pooled-panel count data applications, we observe independent pairs of observations (y_i, x_i), $i = 1, \ldots, n$, and the interest usually centers on the effect of covariates on the conditional mean $\mathrm{E}(y_i|x_i)$, or some other feature of the conditional distribution of $f(y_i|x_i)$. The standard way of introducing covariates is to let $\lambda_{ij} = \exp(x_i'\beta_j)$, $j = 1, 2$, where x_i denotes the vector of covariates and β the parameter vector. This parameterization ensures positive rates and implies a semi-elasticity interpretation for β. Further it allows to treat exposure T_i, the length of the observation period, as a standard covariate. Incorporating exposure explicitly in the model is necessary if T_i varies between individuals and thus cannot be normalized to one.

5.3.3 Decomposing the mean effect

The FH model (see 5.7) has a standard two-part structure, where the two parts are independent. This gives a straightforward decomposition of the overall effect into an effect at the extensive margin and an effect at the intensive margin:

$$\begin{aligned}
&\frac{\partial \mathrm{E}_{FH}(y;\lambda_1(x),\lambda_2(x))}{\partial x} = \\
&\frac{\partial \Pr(y>0;\lambda_1(x))}{\partial x}\mathrm{E}(y|y>0;\lambda_2(x)) + \frac{\partial \mathrm{E}(y|y>0;\lambda_2(x))}{\partial x}\Pr(y>0;\lambda_1(x)) \qquad (5.10)
\end{aligned}$$

It is useful to think of the extensive margin effect as a participation effect (i.e., whether or not one has seen a doctor at all), whereas the intensive margin effect is the effect for participants. This decomposition is so general that it is non-parametrically identified. The question is, however, whether it can be given a causal interpretation. Note that the extensive margin effect is the change in the probability of participation *times the average outcome of participants*. For example, in the context of demand for doctor visits, this would mean that those who are induced by a change in x to see a doctor at least once during a quarter subsequently behave like the "average" individual regarding follow-up visits. This may overstate the true effect. It seems more reasonable to assume that they have below average follow-up visit. But this would violate the independence assumption for the two parts without which such a decomposition does not have a causal interpretation (e.g. Staub, 2013).

In the SH model, the above decomposition (5.10) still holds if one defines

$$\mathrm{E}_{SH}(y|y>0;\lambda_1(x),\lambda_2(x)) = 1 + \frac{\lambda_2 \mathrm{E}(T-t;\lambda_1)}{\Pr(y>0;\lambda_1)}$$

Alternatively, differentiating (5.6) with respect to x, the SH model implies the following decomposition of the partial derivative of the overall mean:

$$\frac{\partial \mathrm{E}_{SH}(y;\lambda_1(x),\lambda_2(x))}{\partial x} = \frac{\partial \Pr(y>0;\lambda_1(x))}{\partial x} + \lambda_2 \frac{\partial \mathrm{E}(T-t;\lambda_1(x))}{\partial x} + \mathrm{E}(T-t;\lambda_1(x))\frac{\partial \lambda_2(x)}{\partial x}$$

Here, the extensive margin effect is the change in the participation probability, multiplied by one, and hence always smaller than the effect under the standard two-part decomposition. The reason is that the marginal observation does not spend any time in the state 2 process, and hence at the margin gets a weight of $\mathrm{E}(y|y>0, T-t=0) = 1$. Also note, that the intensive margin effect can now be further decomposed into a time effect and a productivity effect.

In the application, we will study the effect of a binary policy indicator variable, and it is therefore more relevant to consider a decomposition based on discrete changes. We also need to clarify what we mean by "causal" effects at the two margins. Let λ_1^1, λ_2^1 denote the parameter values with the policy reform in place, whereas λ_1^0, λ_2^0 are the parameters without reform. The extensive margin population is defined as those who change their participation due to the reform. The intensive margin population are those, who participate regardless of reform. For individuals at the extensive margin, the change depends on the direction of the policy effect. With a positive effect, the expected count increases from 0 to $[1 + \lambda_2^1 \mathrm{E}_{EM}(T - t|\lambda_1^1)]$ where $\mathrm{E}_{EM}(T - t)$ is the expected time spent in state 2 by the extensive margin population. For a negative reform effect, there is a decrease from $[1+\lambda_2^0 \mathrm{E}_{EM}(T-t|\lambda_1^0)]$ to zero. Moreover, the fraction of individuals at the extensive margin equals the fraction of individuals induced to participate by the policy, i.e. $\Pr(EM) = \Pr(Y > 0; \lambda^1) - \Pr(Y > 0; \lambda^0)$. The extensive margin effect (EME) is then

$$EME = [1 + \lambda_2^1 \mathrm{E}_{EM}(T - t|\lambda_1^1)] \times \Pr(EM)$$

Similarly, one can obtain the intensive margin effect (IME) as

$$IME = [\lambda_2^1 \mathrm{E}_{IM}(T - t|\lambda_1^1) - \lambda_2^0 \mathrm{E}_{IM}(T - t|\lambda_1^0)] \times \Pr(IM)$$

and $\mathrm{E}(Y, \lambda^1) - \mathrm{E}(Y, \lambda^0) = IME + EME$. Those who do not participate regardless of policy do not affect the mean and thus can be ignored.

The two populations are defined in terms of potential outcomes, one of which is counterfactual, and the model therefore does not per se identify the expected duration spent in the second state by the two groups. A naive approach would be to assume that they are equal. However, it seems more plausible that $\mathrm{E}_{IM}(T - t_1|\lambda_1^1) > \mathrm{E}_{EM}(T - t_1|\lambda_1^1)$. In the application, we make the assumption that the EM population enters the second state sequentially after the IM population, and derive the corresponding decomposition from the underlying structure of the SH model.

5.3.4 Unobserved heterogeneity

The variation of y in empirical applications is often higher than that implied by a Poisson model, even if λ is allowed to depend on covariates. In the Poisson model, one can account for this "overdispersion" by assuming a gamma distributed parameter λ. Let u be gamma distributed with mean 1 and variance α. If, conditional on u, y follows a Poisson distribution with parameter λu, the unconditional distribution of y is negative binomial (NegBin) with mean λ and variance $\lambda(1+\lambda\alpha)$. The SH model can be extended along the same lines. Let u again denote a gamma distributed individual effect and assume that T is normalized to 1. If u equally affects both rates of the SH model, the conditional probability of observing a count k is $f_{SH}(k;\lambda_1 u, \lambda_2 u)$. Integration over the unobserved u gives the unconditional probability (see Appendix B for a derivation):

$$
\begin{aligned}
&f_{SHG}(k;\lambda_1,\lambda_2,\alpha)\\
&\quad = \int_0^\infty f_{SH}(k;\lambda_1 u,\lambda_2 u)\mathrm{Gamma}(u,\alpha)du\\
&\quad = \begin{cases} (\lambda_1/\alpha+1)^{-\alpha} & \text{for k=0}\\ \dfrac{\lambda_1\lambda_2^{k-1}}{(\lambda_2-\lambda_1)^k}\left(\dfrac{\alpha}{\alpha+\lambda_1}\right)^\alpha \left[1-\sum_{j=0}^{y-1}(1-p)^j\, p^\alpha \dfrac{\Gamma(\alpha+j)}{\Gamma(\alpha)\Gamma(j+1)}\right] & \text{for } k=1,\ 2,\ 3,\ldots, \end{cases}
\end{aligned}
\tag{5.11}
$$

with $p=(\alpha+\lambda_1)/(\alpha+\lambda_2)$ and $\mathrm{Gamma}(u,\alpha)$ denoting the density function of a gamma distributed random variable with mean 1 and variance α. If $\lambda_2>\lambda_1$, the term in squared brackets equals the complementary cumulative distribution function of a NegBin distribution. The mean of the stochastic-NegBin hurdle model is given by

$$
\begin{aligned}
\mathrm{E}(y|\lambda_1,\lambda_2,\alpha) &= \int_0^\infty \lambda_2 u + (1-\lambda_2 u/\lambda_1 u)\,(1-\exp(-\lambda_1 u))\mathrm{Gamma}(u,\alpha)du\\
&= \lambda_2 + (1-\lambda_2/\lambda_1)(1-\mathrm{NegBin}(0;\lambda_1,\alpha))\\
&= \lambda_2\mathrm{E}(1-t,\lambda_1,\alpha)+\Pr(y\neq 0,\lambda_1,\alpha)
\end{aligned}
\tag{5.12}
$$

It preserves the essential structure of the mean of the stochastic Poisson hurdle model, and simplifies to it for $\alpha=0$.

5.3.5 Estimation and testing

One can estimate β_1, β_2 and α by maximum likelihood (Stata code is available from the authors upon request). In empirical applications, the interest is often in testing for the presence of excess zeros. Under the null of no additional zeros, $\lambda_1 = \lambda_2$ which implies $\beta_1 = \beta_2$, and the SH model simplifies to a Poisson model. The null hypothesis can therefore be tested by a likelihood ratio test. A similar test is possible for the FH model, whereas the standard ZIP model does not nest the Poisson model (in standard parameterization, the ZIP model converges to the Poisson model if β_1 goes to minus infinity).

Since the SH and FH model are not nested, one has to use a Vuong-Test (Vuong, 1989) to discriminate between them. The Vuong-Test can also be used to compare the SH model with the ZIP. The test is based on the likelihood values and determines whether one of the two models significantly outperforms the other in terms of minimizing the Kullback-Leibler distance. Alternatively, one can select the best model based on an information criterion. However, since the three models have an identical number of parameters, an adjustment for degree of freedom is not needed, and the model with the highest empirical likelihood value is the best choice in terms of Kullback-Leibler distance.

5.4 Application: The 1997 German health care reform and the number of doctor visits

We apply the new stochastic hurdle model in an analysis of the effect of the 1997 German health care reform on the number of doctor visits using data from Winkelmann (2004). In Germany, most of the health cost is paid for by the federal social insurance system. However, there is a co-payment for prescription drugs the amount of which increased substantially in 1997. Winkelmann (2004) used fixed-hurdle and other two-part models to investigate if, and by how much, the demand for doctor visits was affected by this reform, using annual panel data from the German Socio Economic Panel (GSOEP) for the period 1995 to 1999. A key result of that earlier study was that the reform effect was unevenly distributed between the extensive and intensive margins.

The demand for health care is measured by the number of doctor visits in the three months prior to the interview. To estimate the effect, data are pooled and the regression-adjusted difference in the number of visits in the year before the reform, 1996, and the year following the reform, 1998, is taken as a measure for the reform effect. Standard errors are adjusted to account for clustering. Additional controls include: age and age squared, a dummy for male, years of education, a dummy for being actively engaged in sports, two dummies for self assessed health status, log of monthly gross income, a dummy for being married, number of people living in household, a dummy for full time and part time working, a dummy for being unemployed and one for receiving welfare payments, and three quarterly dummies for the timing of the interview (see Winkelmann, 2004, for further details).

Table 5.1 presents the estimation results of the different models. The first column contains the results of a simple Poisson regression. Columns 2 and 3 show the results of the zero-inflated Poisson model (ZIP) where the zero-inflation parameter ϕ is modelled in logit form: $\phi(x, \beta_1) = \exp(x'\beta_1)/(1 + \exp(x'\beta_1))$. A positive coefficient β_1^k indicates that a ceteris paribus increase in the associated regressor x^k increases the probability of an extra zero.

The next four columns show the results for the fixed hurdle model (FH) and the stochastic hurdle model (SH), respectively. Here, the interpretation of the β_1 vector is different, as a positive coefficient means that an increase in the associated regressor increases the rate of the first process, thereby *reducing* the probability of a zero. The zero/positive dichotomy is effectively modelled as a binary model with complementary log-log link. As a consequence, we would expect that the signs of coefficients displayed in columns 4 and 6 are typically opposite to that of column 2, which is indeed the case.

Regarding data fit, the SH model clearly outperforms the other Poisson generalizations in terms of the likelihood value. The comparison based on the likelihood is meaningful as the three models have the same number of parameters. The absence of a stochastic hurdle, i.e., $\beta_1 = \beta_2$, which implies the validity of the simple Poisson model and therefore absence of extra zeros, can be tested by a likelihood ratio test. The hypothesis is rejected under both alternatives, the FH and the SH models.

The qualitative results of the estimated models are similar. The predicted probability of never going to a doctor, for example, is lower for men compared to women with similar characteristics in all models. As expected, having good health increases the probability of no visit and reduces the expected number of visits in the positive part of the model. However, the parameters identify different effects in the three models and there are therefore some systematic differences. For example, the zero part coefficients are similar in the two hurdle models, whereas they are bigger in absolute value in the ZIP model. This is consistent with the different scaling underlying the logit and the complementary log-log models of zeroes. Also, the conditional on positives parameters tend the similar in the ZIP and FH model, but they are larger in absolute value in the SH model. The SH estimates thereby compensate for the fact that the second, or state 2, rate applies only for a fraction of the entire period, in contrast to the FH and ZIP models, where such a time effect is ruled out. It should also be noted that the standard errors of $\hat{\beta}_2$ are substantially higher in the SH model than in FH or ZIP. The dynamic interdependence between the two rates in the SH model means that the second rate cannot be estimated independently of the first rate, reducing the precision of the estimator. On the other hand, $\hat{\beta}_1$ is estimated somewhat more precisely in the SH model than in the FH model (and considerably more so than in the ZIP model), as it uses the distribution of the positives as an additional source of identification.

Regarding the reform effect, all four models find a negative effect on the number of doctor visits in both parts of the model, since $\hat{\beta}_{96} - \hat{\beta}_{98}$ is positive in the logit-part and negative in the remaining models. One key objective of Winkelmann (2004) was to establish whether the reform had different effects in different parts of the outcome distribution, and especially at the extensive margin as compared to the intensive margin. This question is easily answered in the stochastic hurdle model: The estimated effect of the health reform on the first rate is a decrease by $(1-\exp(-0.027-0.071)) \cdot 100\% = 9.3\%$ and is thus larger than the effect on the second rate (a reduction of 5%).

These changes translate into an overall reduction in the mean of 9.86%, which can be decomposed, as in (5.10), into a 5.65 percent reduction of the probability of a positive number of visits, and a 4.46 percent reduction in the average number of visits among participants. These numbers are similar to the ones for the FH and ZIP models (see also Winkelmann, 2004). However, as argued in Section 3.3, this decomposition likely overstates the magnitude of the extensive margin effect, as it assumes that those who are induced by the reform to switch their participation status spend on average the same time in the second state as those who participate regardless of policy reform. If we assume, by contrast, that they have a lower than average mean, because they are the first to switch from participation to non-participation and therefore spend least time in the state 2 process, the extensive margin effect is reduced to -2.02%.

Table 5.2 adds the results of a NegBin model and the stochastic hurdle model with heterogeneity. The estimated dispersion parameter $\hat{\alpha} = 0.163$ is statistically significant, and the heterogeneity model is thus an improvement over the model without unobserved heterogeneity. Nevertheless, the substantive conclusions do not change much. The effect of the reform is again larger for the first rate (-22% vs. -4%) and the overall effect on the number of visits is estimated to be a drop by -9.4%. Interestingly, once we allow for unobserved heterogeneity, the standard decomposition gives an extensive margin effect of -2.48% and extensive margin effect of -7.01%. These estimates are quite close to the causal decomposition of the effect with heterogeneous participants and non-participants in the Poisson SH model.

5.5 Conclusion

Standard models for count data with excess zeros, the zero-inflated and the fixed hurdle model, are not derived from an underlying stochastic process. They are ad-hoc modifications of common count data distributions to allow for a higher proportion of zeros. The lack of an underlying process hampers the interpretation of estimation results. Comparing the parameters of different parts of a model, for example, is difficult. In addition, it is not clear how to deal with varying exposure time in these processes. In contrast, the new stochastic hurdle model is based on a dynamic count process.

We adopted a parsimonious specification, allowing for a one-time change in the underlying rate after the first event. Such occurrence dependence offers a natural way to model a distribution with too many zeros. It leads to a model with the same number of parameters, two sets of coefficients for each regressor, as its main contenders, the zero-inflated count data model and the hurdle-at-zero count model. The usefulness of the model is illustrated by an application on the effect of the 1997 health reform in Germany. We find that the health reform reduced the number of doctor visits. Furthermore, we show that the decomposition into extensive and intensive margin effects is sensitive to the introduction of heterogeneity. Ignoring heterogeneity overstates the effect for those who stop seeing a doctor because of the reform since these persons have a below-average health care utilization anyway.

References

Cameron, A. Colin, and Pravin K. Trivedi (1986): "Economic models based on count data: Comparisons and applications of some estimators and tests", *Jounral of Applied Econometrics*, **1**, 29-53.

Chang, Fwu-Ranq, and Pravin K. Trivedi (2003): "Economics of Self-Medication: Theory and Evidence", *Health Economics*, **12**, 721-739.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship", *Econometrica*, **52**, 909-938.

Heckman, James J., and George J. Borjas (1980): "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence", *Economica*, **47**, 247-283.

Feller, William (1977): *An Introduction to Probability Theory and Its Applications*, second edition, New York: John Wiley & Sons.

Faddy, M. (1997): "Extended Poisson process modelling and analysis of count data", *Biometrical Journal*, **39**, 431-440.

Janardan, Konanur G. (1980): "A Stochastic Model for the Study of Oviposition Evolution of the Pest Callosobruchus Maculatus on Mung Beans, Phaseolus aureus", *Math. Biosciences*, **50**, 231-238.

Lambert, Diane (1992): "Zero-inflated Poisson regression with an application to defects in manufacturing", *Technometrics*, **34**, 1-14.

Mullahy, J. (1986): "Specification and testing in some modified count data models", *Journal of Econometrics*, 33, 341-365.

McShane, B., M. Adrian, E. T. Bradlow, and P. S. Fader (2008): "Count Models Based on Weibull Interarrival Times", *Journal of Business & Economic Statistics*, **26:3**, 369-378.

Pizer, Steven D., and Julia C. Prentice (2011): "Time Is Money: Outpatient Waiting Times and Health Insurance Choices of Elderly Veterans in the United States", *Journal of Health Economics*, **30**, 626-636.

Pohlmeier, Winfried, and Volker Ulrich, (1995): "An Econometric Model of the Two-Part Decisionmaking Process in the Demand for Health Care", *Journal of Human Resources*, **30(2)**, 339-361.

Sari, Nazmi (2009): "Physical Inactivity and its Impact on Healthcare Utilization", *Health Economics*, **18**, 885-901.

Sarma, Sisira, and Wayne Simpson (2006): "A microeconometric analysis of Canadian health care utilization", *Health Economics*, **15**, 219-239.

Staub, Kevin (2013): "A causal interpretation of extensive and intensive margin effects in generalized Tobit models", *Review of Economics and Statistics*, forthcoming.

Street, Andrew, Andrew Jones, and Aya Furuta (1999): "Cost sharing and pharmaceutical utilisation and expenditure in Russia", *Journal of Health Economics*, **18**, 459-472.

Vuong, Quang H. (1989): "Likelihood Ratio Tests for Model Selection and non-nested Hypotheses", *Econometrica*, **57(2)**, 307-333.

Winkelmann, Rainer (1995): "Duration Dependence and Dispersion in Count-Data Models", *Journal of Business & Economic Statistics*, 467-74.

Winkelmann, Rainer (2004): "Health Care Reform and the Number of Doctor Visits – An Econometric Analysis", *Journal of Applied Econometrics*, **19**, 455-472.

Winkelmann, Rainer (2008), *Econometric Analysis of Count Data*, fifth edition, Berlin: Springer.

Yen, Stephen T., Chao-Hsiun Tang, and Shew-Jiuan B. Su (2001): "Demand for Traditional Medicine in Taiwan: A Mixed Gaussian-Poisson Model Approach", *Health Economics*, **10**, 221-232.

A Derivation of the probability function of the stochastic hurdle model

The probability of a zero in the SH model equals the probability of a zero in a Poisson model with rate λ_1. If, $\lambda_1 = \lambda_2$ the SH model degenerates to a Poisson model. For $k = 1, 2, 3, \ldots$ and $\lambda_1 \neq \lambda_2$:

$$\begin{aligned}
&\Pr(Y = k|\lambda_1, \lambda_2) \\
&\quad = \int_0^T \exp(-\lambda_1 t)\lambda_1 \exp(-\lambda_2(T-t))(\lambda_2(T-t))^{k-1}/(k-1)!dt \\
&\quad = \lambda_1\lambda_2^{k-1}\exp(-\lambda_2)\int_0^T \frac{\exp(\lambda_2-\lambda_1)^t(T-t)^{k-1}}{(\lambda_2-\lambda_1)(k-1)!}dt \\
&\quad = \lambda_1\lambda_2^{k-1}\exp(-\lambda_2)\left(\frac{\exp(\lambda_2-\lambda_1)^t(T-t)^{k-1}}{(\lambda_2-\lambda_1)(k-1)!} + \int_0^T \frac{\exp(\lambda_2-\lambda_1)^t(T-t)^{k-2}}{(\lambda_2-\lambda_1)(k-2)!}dt\right) \\
&\quad = \frac{\lambda_2}{\lambda_2-\lambda_1}\Pr(Y = k-1|\lambda_1,\lambda_2) - \frac{\lambda_1\lambda_2^{k-1}}{\lambda_2-\lambda_1}\exp(-\lambda_2)/(k-1)!
\end{aligned}$$

Setting $T = 1$, and solving the recursive formulation of the form $p_k = \alpha p_{k-1} + c_k$ leads to:

$$\begin{aligned}
&\Pr(Y = k|\lambda_1, \lambda_2) \\
&\quad = \alpha^{k-1}\Pr(k = 1|\lambda_1,\lambda_2) + \sum_{j=0}^{k-2}\alpha^j c_{k-j} \\
&\quad = \frac{\lambda_1\lambda_2^{k-1}}{(\lambda_2-\lambda_1)^k}(\exp(-\lambda_1) - \exp(-\lambda_2)) - \sum_{j=0}^{k-2}\left(\frac{\lambda_2}{\lambda_2-\lambda_1}\right)^j \frac{\lambda_1\lambda_2^{k-j-1}}{\lambda_2-\lambda_1}\frac{\exp(\lambda_2)}{(k-j-1)!} \\
&\quad = \frac{\lambda_1\lambda_2^{k-1}}{(\lambda_2-\lambda_1)^k}\left(\exp(-\lambda_1) - \exp(-\lambda_2)\sum_{j=0}^{k-1}\frac{(\lambda_2-\lambda_1)^j}{j!}\right) \\
&\quad = \frac{\lambda_1\lambda_2^{k-1}\exp(-\lambda_1)}{(\lambda_2-\lambda_1)^k}\left(1 - \sum_{j=0}^{k-1}\frac{\exp(-(\lambda_2-\lambda_1))(\lambda_2-\lambda_1)^j}{j!}\right)
\end{aligned}$$

See Janardan (1980) for an alternative derivation.

B Probability function of the SH model with unobserved heterogeneity

This distribution can be seen as a generalization of the negative binomial distribution since they are equal if $\lambda_1 = \lambda_2$. The probability function of the NegBin($y = k, \lambda, \alpha$) distribution is given by

$$\Pr(Y = k|\lambda, \alpha) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\Gamma(k+1)} \left(\frac{\alpha}{\lambda + \alpha}\right)^{\alpha} \left(\frac{\lambda}{\lambda + \alpha}\right)^{k} \qquad \text{for } k = 0, 1, 2, \ldots$$

The probability of a zero in the SH model with unobserved heterogeneity is equal to the probability of a zero in the NegBin model

$$\begin{aligned} \Pr(Y = 0|\lambda_1, \lambda_2, \alpha) &= \int_0^{\infty} f_{SH}(0, \lambda_1 u, \lambda_2 u)\text{Gamma}(u, \alpha) du \\ &= \text{NegBin}(Y = 0, \lambda_1, \alpha) \end{aligned}$$

where f_{SH} is the probability function of the SH model without heterogeneity and Gamma(u, α) denotes the density function of a Gamma distributed random variable u with expectation 1 and variance α. For $\lambda_1 \neq \lambda_2$ and $k = 1, 2, 3, \ldots$

$$\begin{aligned} &\Pr(Y = k|\lambda_1, \lambda_2, \alpha) \\ &\quad = \int_0^{\infty} f_{SH}(k, \lambda_1 u, \lambda_2 u)\text{Gamma}(u, \alpha) du \\ &\quad = \int_0^{\infty} \frac{\lambda_1 \lambda_2^{y-1}}{(\lambda_2 - \lambda_1)^y} \left(e^{-\lambda_1 u} - e^{-\lambda_2 u} \sum_{j=0}^{y-1} \frac{(\lambda_2 u - \lambda_1 u)^j}{j!}\right) \frac{\alpha^{\alpha}}{\Gamma(\alpha)} u^{\alpha-1} e^{-u\alpha} du \\ &\quad = \frac{\lambda_1 \lambda_2^{y-1}}{(\lambda_2 - \lambda_1)^y} \frac{\alpha^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} u^{\alpha-1} e^{-u\alpha} \left(e^{-\lambda_1 u} - e^{-\lambda_2 u} \sum_{j=0}^{y-1} \frac{u^j (\lambda_2 - \lambda_1)^j}{j!}\right) du \\ &\quad = \frac{\lambda_1 \lambda_2^{y-1}}{(\lambda_2 - \lambda_1)^y} \frac{\alpha^{\alpha}}{\Gamma(\alpha)} \left(\int_0^{\infty} u^{\alpha-1} e^{-u(\alpha+\lambda_1)} du - \sum_{j=0}^{y-1} \frac{(\lambda_2 - \lambda_1)^j}{j!} \int_0^{\infty} u^{\alpha+j-1} e^{-u(\alpha+\lambda_2)} du\right) \end{aligned}$$

Using $\int_0^\infty e^{-u\beta} u^{\alpha-1} du = \beta^{-\alpha}\Gamma(\alpha)$ leads to

$$
\begin{aligned}
&\Pr(Y = k|\lambda_1, \lambda_2, \alpha) \\
&\quad = \frac{\lambda_1 \lambda_2^{y-1}}{(\lambda_2 - \lambda_1)^y} \frac{\alpha^\alpha}{\Gamma(\alpha)} \left((\alpha + \lambda_1)^{-\alpha}\Gamma(\alpha) - \sum_{j=0}^{y-1} \frac{(\lambda_2 - \lambda_1)^j}{j!} (\alpha + \lambda_2)^{-(\alpha+j)}\Gamma(\alpha + j) \right) \\
&\quad = \frac{\lambda_1 \lambda_2^{y-1}}{(\lambda_2 - \lambda_1)^y} \alpha^\alpha \left((\alpha + \lambda_1)^{-\alpha} - (\alpha + \lambda_2)^{-\alpha} \sum_{j=0}^{y-1} \frac{(\lambda_2 - \lambda_1)^j}{j!(\alpha + \lambda_2)^j} \frac{\Gamma(\alpha + j)}{\Gamma(\alpha)} \right) \\
&\quad = \frac{\lambda_1 \lambda_2^{y-1}}{(\lambda_2 - \lambda_1)^y} \left(\left(\frac{\alpha}{\alpha + \lambda_1} \right)^\alpha - \left(\frac{\alpha}{\alpha + \lambda_2} \right)^\alpha \sum_{j=0}^{y-1} \left(\frac{\lambda_2 - \lambda_1}{\alpha + \lambda_2} \right)^j \frac{\Gamma(\alpha + j)}{\Gamma(\alpha)\Gamma(j + 1)} \right) \\
&\quad = \frac{\lambda_1 \lambda_2^{y-1}}{(\lambda_2 - \lambda_1)^y} \left(\frac{\alpha}{\alpha + \lambda_1} \right)^\alpha \left(1 - \sum_{j=0}^{y-1} \left(1 - \frac{\alpha + \lambda_1}{\alpha + \lambda_2} \right)^j \left(\frac{\alpha + \lambda_1}{\alpha + \lambda_2} \right)^\alpha \frac{\Gamma(\alpha + j)}{\Gamma(\alpha)\Gamma(j + 1)} \right)
\end{aligned}
$$

Figure 5.1: Poisson, SH, FH and ZIP probability functions $(\mathrm{E}(Y) = 2.35)$

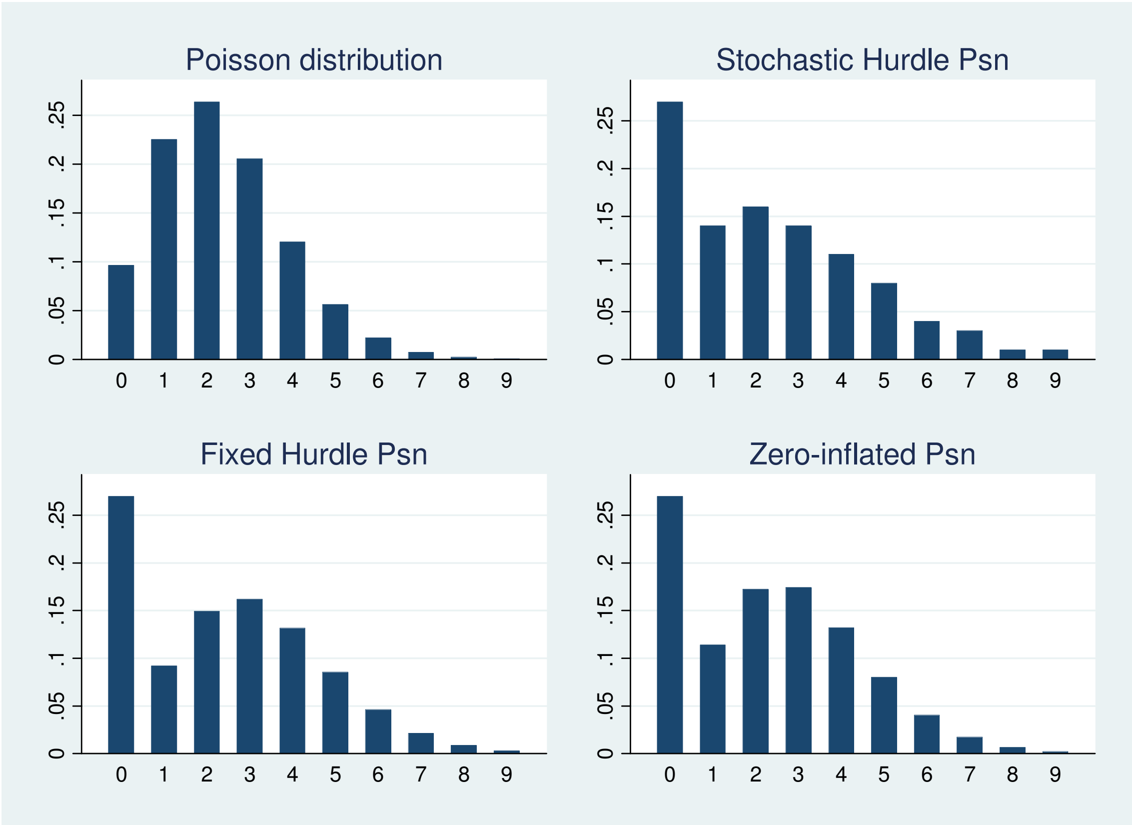


Table 5.1: Excess zero models for the number of doctor visits

	Poisson	Zero-Inflated logit	Zero-Inflated Poisson	Fixed Hurdle λ_1	Fixed Hurdle λ_2	Stochastic Hurdle λ_1	Stochastic Hurdle λ_2
Age$\times 10^{-1}$	-0.106	0.510**	-0.007	-0.262**	-0.004	-0.267**	0.060
	(0.086)	(0.134)	(0.087)	(0.069)	(0.087)	(0.065)	(0.144)
Age$^2 \times 10^{-3}$	0.158	-0.658**	0.029	0.337**	0.026	0.327**	-0.043
	(0.105)	(0.165)	(0.106)	(0.085)	(0.106)	(0.079)	(0.175)
Male	-0.209**	0.715**	-0.055*	-0.388**	-0.052	-0.364**	0.010
	(0.028)	(0.042)	(0.027)	(0.022)	(0.027)	(0.020)	(0.046)
Educ$\times 10^{-1}$	-0.058	-0.400**	-0.161**	0.169**	-0.161**	0.192**	-0.302**
	(0.047)	(0.084)	(0.048)	(0.042)	(0.048)	(0.039)	(0.078)
Sport	0.047*	-0.278**	-0.015	0.141**	-0.016	0.135**	-0.065
	(0.022)	(0.039)	(0.022)	(0.020)	(0.023)	(0.019)	(0.038)
Goodh	-0.611**	0.612**	-0.428**	-0.460**	-0.429**	-0.385**	-0.533**
	(0.021)	(0.038)	(0.021)	(0.019)	(0.021)	(0.019)	(0.036)
Badh	0.813**	-0.939**	0.653**	0.489**	0.653**	0.194**	0.928**
	(0.026)	(0.061)	(0.024)	(0.027)	(0.024)	(0.024)	(0.049)
Loginc	0.093**	-0.225**	0.033	0.132**	0.035	0.111**	0.005
	(0.028)	(0.047)	(0.028)	(0.024)	(0.028)	(0.023)	(0.050)
Year=96	0.001	-0.138**	-0.027	0.063**	-0.028	0.071**	-0.077
	(0.023)	(0.042)	(0.024)	(0.021)	(0.024)	(0.020)	(0.045)
Year=97	-0.030	-0.017	-0.033	0.002	-0.033	0.022	-0.084
	(0.024)	(0.041)	(0.025)	(0.021)	(0.025)	(0.020)	(0.046)
Year=98	-0.105**	0.072	-0.078**	-0.064**	-0.080**	-0.027	-0.128**
	(0.025)	(0.042)	(0.026)	(0.022)	(0.026)	(0.021)	(0.047)
Year=99	-0.099**	-0.076	-0.111**	0.008	-0.113**	0.036	-0.186**
	(0.026)	(0.045)	(0.027)	(0.022)	(0.027)	(0.021)	(0.048)
Constant	0.905**	-0.151	1.515**	-0.212	1.496**	-0.242	2.220**
	(0.262)	(0.431)	(0.266)	(0.218)	(0.267)	(0.205)	(0.461)
log-likelihood	-86,566	-77,980		-77,999		-72,137	

n=32837. Model in addition includes three quarterly dummies, three dummies for the employment status, a dummy for receiving welfare payments, one for being married, and household size. Standard errors in parentheses (pooled model with cluster robust standard errors at the individual level). **, * denote statistical significance at the 1%, 5% significance levels, respectively.

Table 5.2: Models with unobserved heterogeneity

		Stochastic Hurdle	
	NegBin	λ_1	λ_2
Age$\times 10^{-1}$	-0.194*	-0.483**	-0.120
	(0.078)	(0.125)	(0.093)
Age$^2 \times 10^{-3}$	0.265**	0.627**	0.166
	(0.095)	(0.154)	(0.112)
Male	-0.291**	-0.725**	-0.107**
	(0.025)	(0.041)	(0.029)
Educ$\times 10^{-1}$	-0.042	0.276**	-0.185**
	(0.045)	(0.069)	(0.052)
Sport	0.074**	0.251**	-0.001
	(0.023)	(0.035)	(0.027)
Goodh	-0.626**	-0.806**	-0.549**
	(0.021)	(0.035)	(0.026)
Badh	0.826**	1.081**	0.805**
	(0.027)	(0.057)	(0.029)
Loginc	0.085**	0.223**	0.030
	(0.028)	(0.044)	(0.033)
Year=96	0.007	0.118**	-0.035
	(0.022)	(0.039)	(0.028)
Year=97	-0.035	0.002	-0.043
	(0.023)	(0.037)	(0.029)
Year=98	-0.086**	-0.101**	-0.074*
	(0.024)	(0.039)	(0.030)
Year=99	-0.079**	0.016	-0.118**
	(0.025)	(0.040)	(0.031)
Constant	1.106**	0.414	1.519**
	(0.252)	(0.406)	(0.303)
ln(α)	0.097**	0.153**	
	(0.013)	(0.029)	
log-likelihood	-64,612	-64,267	

n=32837. Model in addition includes three quarterly dummies, three dummies for the employment status, a dummy for receiving welfare payments, one for being married, and household size. Standard errors in parentheses (pooled model with cluster robust standard errors at the individual level). **, * denote statistical significance at the 1%, 5% significance levels, respectively.